Journal of the
OPTICAL SOCIETY
of AMERICA

VOLUME 39, NUMBER 2 FEBRUARY, 1949

EDITORIAL COMMENT

Publication Costs and a New Format

READERS will notice a change in page format in the current issue. Though articles are included which had been set up before the change was made, in the format with which our readers have become familiar over the past fifteen years, the remaining pages of the Journal take on a new appearance. Though the type used is of the same size as before, narrower margins and spacing between lines have been adopted as standard for succeeding issues, changes which permit putting approximately 20 percent more reading matter on each page. The Editors have reluctantly adopted this more crowded page because of rapidly increasing costs of publication. This is probably only the first of a series of changes dictated by the necessity of introducing economies in publication wherever possible. Similar changes in format are being adopted for the same reason by several other physics journals.

Costs of publication, as of almost everything else, have been rising rapidly since the war, and a recent additional large jump in printing costs has made necessary either a reduction in the quantity of research results printed, a large increase in Journal income, or the adoption of economy measures which, while necessarily diminishing the attractive appearance of the Journal pages, would reduce the cost of publishing a given quantity of research material. The first method is impracticable, since the quantity of important research material to be published is continually increasing.

The Editors have adopted as a firm policy the prompt publication, as nearly as possible within the minimum three months required for editorial and publication mechanics, of all material considered to be of adequate quality. Consideration has been given to the desirability of requiring contributors to condense their articles by some considerable percentage, as is the custom with publications in certain sister sciences. It has been felt that this would not be wise at the present time, and that the strain from necessary increased costs should be distributed as evenly as possible over all sources of Journal income.

The Directors of the Optical Society therefore undertook at their annual meeting in October (see Minutes in December, 1948, issue, page 1107) to raise all subscription rates and dues in the various membership categories by a small amount. This appears entirely justified, not only on the basis that the subscriber's dollar is now worth much less than it was before the war, but also because of the greatly increased amount of material being published in each annual volume (1135 pages for 1948 as compared with 508 pages for 1938).

The possibility of increasing income by the acceptance of a few especially pertinent and selected advertisements is now being investigated as a possible source of additional income. In the meantime, it is necessary for the Board of Editors to exercise even more editorial vigilance than usual to insure that all material published be as succinct as possible. All in all, the Society has managed matters so efficiently that the cost of the Journal to members has for many years been considerably less than that of comparable journals. The increase in dues comes, as of next year, because of circumstances over which the Society has no control.

Sunlight and Skylight as Determinants of Photographic Exposure. II. Scene Structure, Directional Index, Photographic Efficiency of Daylight, Safety Factors, and Evaluation of Camera Exposure

LOYD A. JONES AND H. R. CONDIT
Communication No. 1155-II from the Kodak Research Laboratories, Eastman Kodak Company, Rochester, New York
(Received September 7, 1948)

Following a brief review of Part I of this communication, additional factors involved in the computation of correct camera exposure are discussed. The first of these is the structure of the scene. Two aspects of structure, namely, *spatial structure* and *local illumination* conditions, are recognized. Four categories of spatial structure, definable, at least in a qualitative manner, in terms of distance between scene elements and camera, are described and illustrated pictorially. Local illumination conditions are defined in terms of the extent of shading and four definitely recognizable aspects are described.

The minimum scene luminance, B_s min., is determined to a considerable extent by aspects of scene structure, and quantitative values corresponding to combinations of structural elements are given.

When an appreciable part of the light entering the scene space comes from a source which subtends a small solid angle, the minimum scene luminance is dependent to some extent upon the direction, relative to the optical axis of the camera system, from which this unidirectional illumination enters the scene space. The angular relationships between this component of the incident light and the optical axis of the camera which results in conditions defined as *front lighting*, *side* and *top lighting*, and *back lighting* are defined and relative values of B_s min. are given corresponding to these three directional lighting conditions.

The speed of photographic negative materials is defined in terms of radiant energy having a specified spectral composition which has been standardized by international agreement. The spectral composition of the radiant energy which, *on the average*, is reflected from the scene elements of minimum luminance is not identical to that used in the measurement of speed. It is therefore necessary to include a term in the formula for the computation of correct camera exposure, the magnitude of which depends upon the difference in the spectral qualities of the two radiant energies. This factor is expressed in terms of the photographic efficiency of the radiant energy reflected *on the average* from the scene elements of minimum luminance. This efficiency can be computed, if certain functional relations are known, or it may be measured by methods of photographic photometry. Evidence shows that the values obtained by the two methods are in fairly good agreement. However, it seems advisable to use a value derived by the direct methods of photographic photometry. In this connection the spectral composition of radiant energy in sunlight and skylight is considered in detail. The use of the term *color temperature* as a significant specification for the spectral composition of radiant energy emitted by natural and artificial sources is examined critically and the conclusion is reached that in the majority of cases a color temperature specification is not adequate for defining the quality, i.e., spectral composition, from the standpoint of photographic efficiency.

As a final step in the derivation of the formula for correct camera exposure, certain safety factors are considered and recommended. Up to this point the treatment has been analytical, involving the consideration of many factors. To make the information of practical use, the treatment must be integrated and simplified in such a way that values of correct camera exposure can be determined simply and expeditiously. Three possible methods are discussed for this purpose: a purely arithmetic, a purely logarithmic, and a combination of arithmetic, logarithmic, and simple slide rule operation. The last of these possible techniques turns out to be the simplest and probably the most easily applied in practical work.

OUTLINE

VII. Scene Structure
- A. Spatial Structure
  1. Distant scenes
  2. Semidistant scenes
  3. Near scenes
  4. Close scenes
- B. B_s min. and Viewing Distance
- C. Local Illumination
  1. Unshaded
  2. Light shade
  3. Heavy shade
  4. Dense shade
- D. Relation between B_s min. and Scene Structure

VIII. Directional Lighting
- A. Front Lighting
- B. Side and Top Lighting
- C. Back Lighting

IX. Photographic Efficiency of Radiant Energy
- A. Spectral Composition of Radiant Energy Coming from the Scene Elements of Minimum Luminance
- B. Sensitometric Measurement of the Relative Photographic Efficiency of Radiant Energy
- C. Spectral Composition of Various Phases of Sunlight and Skylight
- D. Color Temperature, T_c
  1. Various phases of sunlight and skylight
  2. Artificial sources of radiant energy, selective radiators
  3. Artificial sources approximating non-selective radiators

X. Safety Factors

XI. Determination of Correct Camera Exposure
- A. Arithmetic Method
- B. Logarithmic Method
- C. Combined Arithmetic, Logarithmic, and Slide Rule Method.

XII. Conclusion

INTRODUCTION

IN the first part of this communication [J. Opt. Soc. Am. **38**, 123, (1948)], the relation between the *illuminance* incident on the negative material in the camera and the luminance of the corresponding scene element in the object was discussed at some length, and Eq. (3) showing the relation between these two factors was given. This equation ignores the non-image-forming *flare light* which is usually

present to a greater or lesser extent under all conditions in practical photography. The influence of this flare light upon the illuminance distribution over the surface of the negative material was discussed and a value based upon the statistical average of a large number of observations under a wide variety of conditions in exterior scenes was given. By introducing the flare factor and combining Eq. (3) with Eq. (14c) (the equation for camera exposure) and with Eq. (19) (which defines the speed of a photographic material), Eq. (20a) was obtained. This gives the value of correct camera exposure, t/f^2, which is the least camera exposure that will yield a negative from which a print of excellent quality can be made.

The voluminous data available in the literature concerning the amount of radiant energy and light reaching the earth's surface from the sun and sky were reviewed and summarized. A large proportion of these data was derived from the work of Abbot and Fowle and their co-workers at the Smithsonian Institution, and from the contributions of Kimball and his associates of the U. S. Weather Bureau. Other data from sources in this country and abroad were also used in obtaining the final averages which, in our opinion, represent the most reliable evaluation of the various phases of sunlight and skylight. In this work use was made of Moon's very excellent summary of the direct solar radiant energy and light available at the earth's surface for different values of air mass.

There are several possible methods of evaluating the amount of light reaching the earth's surface and it is quite obvious that these different methods may not be of equal significance for the purposes of predicting correct camera exposure. The relative merits of using an evaluation on the *normal*, *perpendicular*, and *horizontal* planes were analyzed. A relatively new concept in the evaluation of the amount of light arriving at the earth's surface was introduced. This is luminous density, where luminous density is used as an abbreviation of the longer but more specific phrase, *volume density of luminous energy*. An evaluation made in this way is independent of the direction from which the light reaches the earth's surface and, according to our analysis of the problem, is more significant as an index of correct camera exposure than any of the previously mentioned three methods, all of which depend upon directional aspects.

It was concluded from an examination of a considerable body of data that the most significant characteristic of a scene for the purpose of determining correct camera exposure is the *minimum object luminance*, B_0 min.

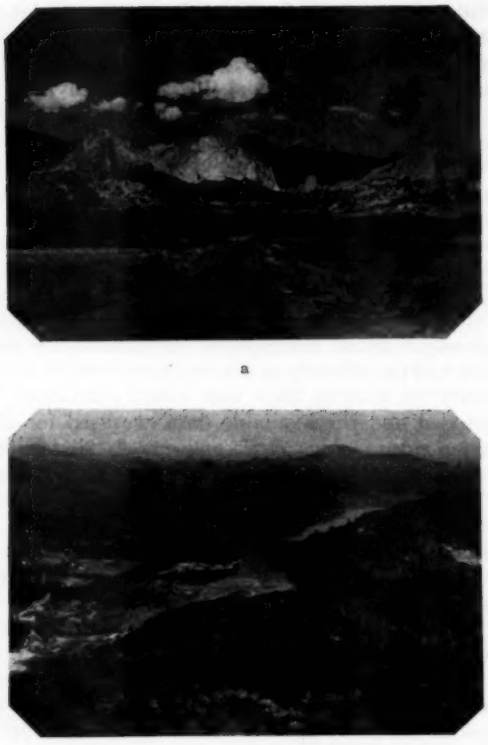
The luminous density at the earth's surface, assuming a clear atmosphere, is determined by solar altitude. Hence, in order to evaluate luminous density for a particular condition, a knowledge of the solar altitude is essential. While under certain conditions it is possible to measure solar altitude by relatively simple methods, there are many conditions under which this is not possible. Solar altitude, however, can be computed for any *geographical latitude* and for a specified *hour of the day* (before or after local solar noon) and for any *month of the year*. This has been done and the relative values of luminous density (in the form of *light indices*, L_v) are shown in tabulated form.

The value of luminous density resulting from light *arriving* at the earth's surface is, of course, dependent upon atmospheric conditions. A discussion of the effect upon the magnitude of luminous density of various haze and cloud conditions is discussed and the available data are used for the assignment of *atmospheric indices*, A, to various atmospheric conditions which can be described with sufficient precision so as to be recognizable with a fair degree of certainty.

Before the correct camera exposure can be completely evaluated, it is necessary to determine the effects of certain other variables. Among these may be mentioned the *structure* of the scene being photographed, of which two distinct aspects must be recognized, namely, the spatial *structure* and *local illumination conditions*. It is also necessary to consider the effect upon the minimum object luminance, B_0 min., of certain conditions where a relatively large proportion of the total light entering the scene space is approximately unidirectional, that is, coming from the completely unobscured sun or the sun obscured by light or medium haze conditions. Furthermore, the *photographic efficiency* of the radiant energy reflected from the scene elements of minimum luminance relative to that of the radiant energy used in the measurement of the speed of photographic negative materials must be considered. Also, it seems advisable for practical purposes to include certain *safety factors* in the final equation for the computation of correct camera exposure. These aspects of the problem will be discussed in this present communication. Finally, the data thus obtained will be simplified and integrated so they can be used simply and expeditiously in the determination of the correct camera exposure for any particular set of conditions.

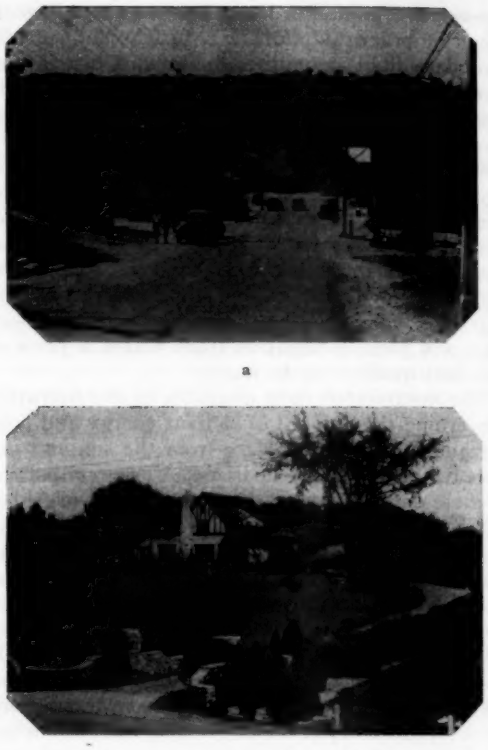
VII. SCENE STRUCTURE

The term *scene structure*, as used in this discussion, includes the significant characteristics of a scene which depend upon the location of its elements in the three-dimensional space in which they are distributed, and upon certain local modifications of the general illumination which are attributable to the position, size, and character of these elements.



a b

FIG. 24. Examples of distant scenes.



a b

FIG. 25. Examples of semidistant scenes.

There are two aspects of scene structure which may be designated as *spatial structure* and *local illumination*.

We consider that the aspect of scene luminance which is the most significant determinant of correct camera exposure is the *minimum* luminance, B_o min. Those structural characteristics of the scene which determine B_o min. are therefore of greatest importance for the present purpose.

A. Spatial Structure

Since the scene elements are distributed in three-dimensional space, their disposal in each of three directions must be indicated for a complete specification of the structure. A careful analysis of the problem reveals no valid reason for concluding that the horizontal distribution (to the left or right of the line of sight) of the scene elements exerts any systematic control on B_o min., nor, in fact, can any tenable ground be established for a significant correlation between B_o min. and the vertical distribution (above or below the line of sight) of the scene elements. On the other hand, the location of the scene elements in the longitudinal direction, i.e., with respect to the distance between camera and scene elements, does have a systematic and appreciable effect upon B_o min.

An inspection of the available exposure tables, exposure guides, and discussions of the exposure problem reveal that this fact has been recognized, at least implicitly, by almost every investigator of this subject. It is well known that such scenes as "distant landscapes," "open landscapes," "distant mountains," "seascapes," and so forth, may be given less camera exposure than less distant compositions, even when the general lighting conditions are the same, and that camera exposure must be increased more and more as the scene elements are closer and closer to the camera.

Scenes can be classified into four distinct groups according to this distance factor. They are: (a) distant, (b) semidistant, (c) near, and (d) close. With constant values of light index, L_v, and of atmospheric index, A, the spread of B_o min. values within each group is relatively small and the average B_o min. for each group decreases progressively from (a) to (d).[10,*]

These classes cannot be defined in terms of dis-

* Footnotes *a–t*, references 1–53, Tables I–XIV, and Figs. 1–23 may be found in Part I of this paper (J. Opt. Soc. Am. 38, 123 (1948)).

tance expressed in unit of length, such as feet, yards, miles, etc. A more rational unit is the focal length of the camera lens with which the scene is being photographed. Thus, a *close* scene might be defined as one in which the distance from the camera to the most important scene elements is less than a times the focal length, F, of the lens, and a near scene as one in which this distance is greater than aF but less than bF, etc. However, even this method is subject to some objections,[u] and so it seems better to rely entirely on the more qualitative verbal descriptions.

1. Distant scenes.—Compositions classified as *distant* contain few, if any, objects of interest near the camera. The foreground is open and almost devoid of important detail. The scene is composed usually of a multitudinous collection of elements at relatively great distances, and this entire collection of objects merges into the composition of the scene as a whole. Although certain elements or groups of elements of the composition are distinctly separable, attention is not generally centered on any particular one of the objects unless it is very large and occupies an appreciable portion of the entire picture area. We are concerned largely with compositional factors, such as balance of light and shade, and the relationships of the parts to the whole. Some specific cases may be mentioned, such as distant mountains, rivers, lakes, open marine views, etc. Many such scenes include large sky areas with or without cloud formations. This type of scene is illustrated in Fig. 24.

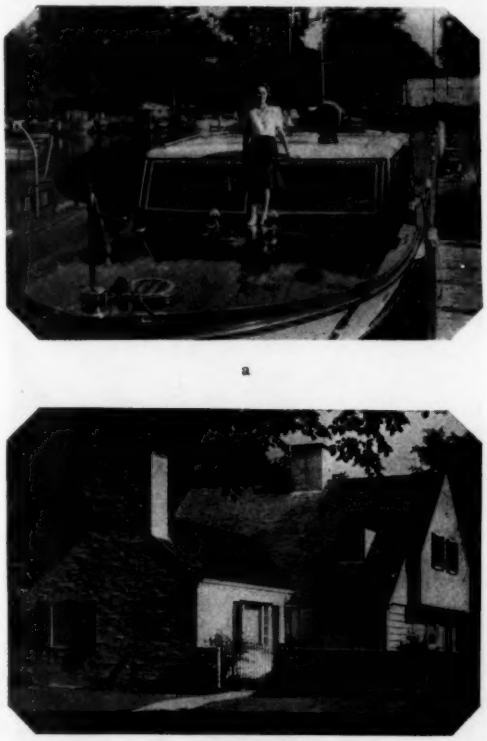
2. Semidistant scenes.—The component parts in semidistant scenes are considerably nearer the camera than in distant scenes. They usually include at least a portion of sky. They may include groups of people who are not close enough to the camera to become objects of chief interest, but are merely incidental to the composition as a whole. The scenes in Fig. 25 are typical of this class.

3. Near scenes.—Near scenes are composed of objects relatively near the camera. Figure 26 shows typical examples. Usually such a composition contains a single object, or group of objects, on which attention is fixed. While there may be many other elements in the composition, they are regarded largely as environmental or background elements. They frequently include small areas of sky but are identified chiefly by the fact that the object on which attention is centered occupies a considerable portion of the picture area. In this group can be placed a large percentage of all the pictures taken by the average amateur. They include full-length portraits of an individual, groups of individuals, pictures of a single or a small group of buildings, a part of a garden, many street scenes, etc.

4. Close[v] *scenes.*—Examples of this subdivision are shown in Fig. 27. Here there is a single object, or very few objects of maximum importance and interest. Everything else in the composition is subordinate and the one object or very few objects are frequently so close to the camera that they occupy almost all of the picture area. Compositions of this type can readily be identified. Specific examples are the head and shoulders of a single individual or of a small group of persons, a single flower or a vase of flowers which occupies the greater part of the composition, a bit of architectural detail, such as a doorway, a window, etc.

B. B_o min. and Viewing Distance

The question of why the minimum object brightness depends upon distance may be raised, especially since a well-known law of optics states that the



a

b

FIG. 26. Examples of near scenes.

[u] Conditions other than the distance expressed in focal lengths of the lens have a bearing on the size of the scene element which must be considered as controlling the effective minimum brightness. Thus, the conditions of printing and of viewing the final production are factors which enter into this problem and which deserve some consideration.

[v] This word denotes practically the same compositional characteristics as the term *close-up* used so frequently in motion picture terminology. Since the addition of the preposition seems superfluous, we prefer the simpler one-word designation.

a

b

FIG. 27. Examples of close scenes.

luminance of a surface is independent of the distance from that surface to the observer. This law assumes that the surface in question is large enough to be seen as a distinct and separable element of the visual field. If the angle subtended by the surface element at the eye of the observer is less than the *resolving power* (or visual acuity, as it is frequently called) of the observer's eye, the surface element is not seen as a separable element but blends into the background. The distance at which this occurs is sometimes referred to as the *blending distance*.

The visual field, only a part of which is usually included in the picture area, is made up of an almost infinite number of surface elements varying in luminance, size, shape, orientation, and disposition. Viewed from the camera position, some of these are invisible to the observer either because they are so small that they are not resolved as separate elements or because the luminance difference between them and their immediate surroundings is below the sensitivity of the human eye to luminance differences.

It is assumed here that the best photographic quality is achieved when the positive reproduces with the greatest fidelity attainable the *appearance* of the visual field as seen by an observer at the camera position. Luminance differences which are invisible to the normal observer need not be reproduced in order to achieve an excellent positive. All that are visible should be reproduced.

The minimum object luminance, as seen from the camera position, may depend therefore upon the distance of the scene elements from the camera position. This is illustrated by Fig. 28, which is a reproduction of a test object made specifically to

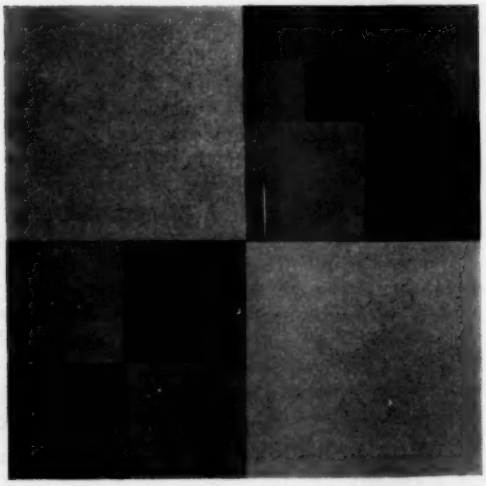


FIG. 28. Conventionalized test object illustrating dependence of effective B_s min. on viewing distance.

demonstrate this point. Let us assume that this chart is introduced into a scene and so located within the structure of the scene that the B_0 min. is that of the small squares designated as 6 in Fig. 29 and that the illumination on the chart is

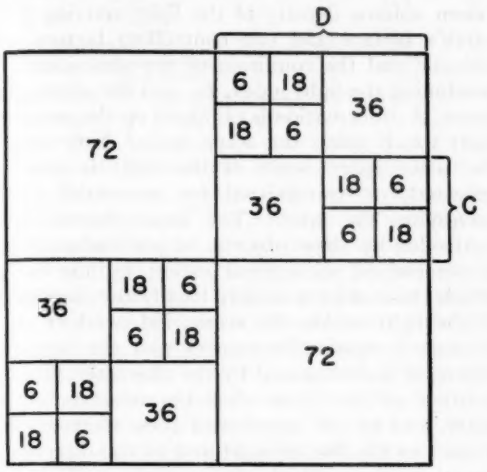


FIG. 29. Luminance values of elements in the test object shown in Fig. 28.

such that the numbers are the values of the luminances of the corresponding areas in Fig. 28. Choosing a point of observation from which all of the elements of the chart are resolved, we find that the B_0 min. for this scene is 6 footlamberts. Now suppose that the observer moves away from the test chart until the small squares designated as 6 and 18 are no longer resolved, but the unit C, made up of these four elements, has the appearance of a uniform dark gray, the luminance of which is the average of its components, namely, 12 footlamberts. This now becomes the effective B_0 min. If the distance of observation is increased still further, a point is reached where the squares which compose the upper right and lower left quadrants, unit D, are all below the resolving power of the eye and hence blend into a uniform gray having a luminance of 24 footlamberts. Thus, again, B_0 min. has been increased. Further increase in the distance of observation will make it impossible to see as separate elements the four quadrants of the test object. When this condition has been attained, the entire test object appears as a gray square with an apparently uniform luminance of 48 foot-lamberts which now becomes the effective B_0 min.

While this test object is a highly simplified and conventional structure, it faithfully represents many conditions met in practice. For instance, Fig. 30 is a reproduction of a photograph of a tree trunk with deeply furrowed bark. Seen at close range, the luminance varies appreciably from point to point and B_0 min. is determined by the luminance of small areas located at the bottoms of the bark fissures. When this same tree trunk is viewed from a much greater distance, possibly three or four hundred feet, it appears as a uniform gray of a luminance much greater than the B_0 min. as observed from the shorter viewing distance.

Hence, the meaning of the term *minimum object luminance*, which we have used so frequently in previous pages, must be qualified or restricted. It cannot be interpreted to mean the actual minimum

FIG. 30. Actual object illustrating dependence of effective B_0 min. on viewing distance.

luminance which can be found if the scene is analyzed into its smallest component elements. This qualification may be expressed by saying that the term B_0 min. refers to the luminance of the darkest scene element in which detail is visible to an observer, having normal visual acuity and normal luminance discrimination, at the camera position. A definite technique must be followed, therefore, in measuring values of B_0 min. if they are to be significant in the determination of correct camera exposure. For instance, a device with an

optical system giving considerable magnification might be employed for measuring this quantity. With such an instrument, the luminance of a scene element far below the resolving power of the unaided human eye might be measured, thus indicating more exposure than is necessary. Instruments for this purpose should be designed to measure scene elements of an angular size approximately equal to the resolving power of the human eye. This is by no means a fixed or constant value, since it depends upon the characteristics of the visual field in which the visual acuity is determined. For instance, visual acuity depends in a great measure upon the luminance contrast between the elements in the pattern and also upon the luminance composition of the field of view. Visual acuity has been studied by many observers and a wealth of data is available on the variation of resolving power for various conditions. Most of the data have been obtained under laboratory conditions and it is difficult to draw definite conclusions from them for use in field work. Moon has summarized the literature of this subject and gives an excellent bibliography.[2] Under favorable conditions of illumination, a black line subtending an angle of 45 seconds at the observer's eye, on a white background, is visible as a line. In this case the contrast, the ratio of background luminance to line luminance, is relatively high, being approximately 25. However, for the luminances which are usually found in the shadow regions of exterior scenes and with the luminance differences which usually produce the visible detail in these regions (contrast of 2 or even less rather than 25), visual acuity is much reduced. This is particularly true in scenes where the visual field includes large areas of high luminance, such as sky or high reflecting surfaces illuminated by sunlight. Such areas act as sources of more or less strong "glare," thus reducing visual acuity and the sensitivity of the eye to luminance differences.

From a study of these data it is evident that no single value of visual acuity can be expected to apply to the wide range of conditions existing in exterior photography. However, since at present we are interested only in the recognition of detail in the shadow regions, it does not seem impossible to arrive at a reasonable average value. It has been found that the minimum luminance varies from about 1 to 100 footlamberts and that, as a rule, the luminance contrast which forms the detail in the shadow regions is quite low, of the order of 2 or less. From observations made with a series of test charts, similar to that shown in Fig. 28, which were actually placed in the shadow areas of a variety of exterior scenes, it has been concluded that the effective acuity of a normal observer for the perception of shadow detail lies between 3 and 7 minutes of arc.*

C. Local Illumination

Thus far we have formulated the relation between volume density of the light arriving at the earth's surface and two controlling factors, solar altitude and the condition of the atmosphere. By combining the light index, L_v, and the atmospheric index, A, information is obtained on the amount of light which *enters* the scene space. After entering the scene space, some of this light is absorbed, reflected, or transmitted by terrestrial objects occupying the space. The *local illumination* is controlled by these objects. In particular, B_c min. is determined to a great extent by the way in which these objects modify locally the distribution of the light within the scene, independent of the distance between the camera and the scene elements. It is determined by the character, size, and position of the scene elements relative to each other, and by the location of these elements with respect to the line of sight and to the direction or directions from which the sunlight and skylight enter the scene. In general, this local illumination factor is of greater importance in scenes of the *close* and *near* types than in distant and semidistant compositions.

At least four distinct conditions of local illumination can be described with sufficient clarity to be easily recognized. They are characterized by the extent to which the scene elements of minimum luminance (in which detail is visible) are shielded from direct sunlight and direct skylight.

1. Unshaded.—This condition is found where the scene as a whole is illuminated by direct sunlight and usually by large areas of sky. This is termed *unshaded*, although some areas of the scene do not receive direct sunlight. In scenes of this class, the illumination of the elements of least luminance in which detail is visible is generally caused by light coming directly from small angular elements of the sky supplemented by some sunlight and skylight which has been diminished in intensity by single or a few multiple reflections from the nearby components of the scene. The highest B_c min. values are usually associated with scenes in this group.

2. Light shade.—Terrestrial objects, such as mountains, hills, trees, buildings, and the like, may be so located in the scene space that practically all of the scene elements are shielded from direct sunlight. The scene as a whole is usually illuminated

* It is interesting to note that the photometer used by us, in the study of brightness distribution in exterior scenes, included within the photometric field an angular distance of 7 minutes of arc and hence B_c min. values obtained with this instrument should approximate quite closely the brightness of the darkest scene element in which detail is visible.

TABLE XV. Values of average minimum luminance for sunlit scenes. Clear atmospheric conditions, $A=0$.

Spatial structure	Local illumination	No. of scenes	Average minimum luminance B_o min. Observed	MB_o min. Computed	MB_o min. Smoothed
Distant	Unshaded	28	55.0	68.8	64
Semidistant	Unshaded	39	24.5	30.6	32
Near	Unshaded	38	12.5	15.6	16
Near	Light shade	12	6.0	7.5	8
Near	Heavy shade	7	3.0	3.75	4
Near	Dense shade	0	1.5*	1.88*	2*

* Extrapolated.

by relatively large solid angular elements of the sky, π-steradians or more. The scene elements of minimum luminance receive no direct sunlight and relatively small amounts of singly or multiply reflected sunlight. The illumination of the shadow areas is a result largely of direct skylight and the scene elements of *minimum luminance* receive light directly from relatively small angular elements of the sky enhanced by skylight reflected from other nearby elements of the scene. The highlights of the scene are usually illuminated by direct sunlight, by large areas of the sky, or by a mixture of the two in various proportions.

3. Heavy shade.—Here the scene elements of minimum luminance receive no direct light from the sun and little, if any, directly from the sky. They are illuminated chiefly by skylight which has been reflected from other nearby scene elements and in some cases by some light coming directly from very small angular elements of the sky. The highlights of the scene may still be illuminated by direct sunlight, though usually they receive light only from moderately large areas of the sky.

4. Dense shade.—Usually in scenes of this type no direct sunlight reaches any of the scene elements. B_o min. is usually associated with scene elements which are almost completely surrounded by nearby numerous objects so that the multiply reflected sunlight and skylight which reach them have been greatly attenuated. The middle tones of such scenes are also illuminated to a great extent by light reflected from adjacent objects rather than by direct sunlight or skylight. The highlights are usually illuminated only by sky areas of small size. Such conditions are frequently encountered in dense woods, in deep and narrow ravines, or well back under a low roof, etc.

The average B_o min. decreases progressively from the first to the last of these four local illumination conditions because of progressively heavier shading of the shadow areas and is usually, but not always, accompanied by an increase in size of the shaded region relative to the total picture area. On the average, the luminance of the highlights also decreases, progressively but not as rapidly as the decrease in B_o min. As a result, the luminance scale of the scene tends to increase progressively from the unshaded to the heavy shaded condition. However, since the highlight of a dense-shade scene seldom is illuminated by direct sunlight, the BS_o of such scenes is, in most cases, lower than for other shaded conditions.

While conditions of light, heavy, and dense shade are found more frequently in scenes in the near and close groups than in the distant and semidistant types, there are cases where it is necessary to consider the local illumination of the distant and semidistant scenes. For example, imagine a hilly or mountainous country with the camera looking down into a deep valley. The sun may be located behind the camera and well to the left or right of the line of sight, but still meet the specification of "front lighting." The shadow cast by the hills at one side strikes out over the valley floor. This shaded area is illuminated by large areas of sky but no direct sunlight. This condition is properly classified as *light shade*. Perhaps the camera is looking into a deep ravine, canyon, or narrow valley with steep slopes on both sides, with a stream of water flowing at the bottom. Unless the sun is almost directly behind the camera, the bottom of these depressions and the lower portions of the side walls will be illuminated only by relatively small solid angular elements of the sky. Such a condition must be recognized as *heavy shade* and the camera exposure adjusted accordingly.

D. Relation between B_o min. and Scene Structure

It is necessary now to establish the relationship between the various aspects of scene structure which have just been discussed and the minimum object luminance, B_o min. The data available for this purpose are relatively meager and for the

TABLE XVI. Luminous density ratios, R_o. Latitude 40°N, clear atmospheric conditions.

	Hours before or after noon, local solar time							
	0	1	2	3	4	5	6	7
June	1.01	1.02	1.03	1.08	1.14	1.31	1.83	8.0
July May	1.02	1.02	1.04	1.09	1.16	1.36	2.06	
Aug. April	1.03	1.04	1.07	1.11	1.23	1.56	3.88	
Sept. March	1.07	1.08	1.11	1.20	1.41	2.33		
Oct. Feb.	1.13	1.15	1.21	1.36	1.86	9.15		
Nov. Jan.	1.24	1.27	1.36	1.68	3.27			
Dec.	1.29	1.33	1.47	1.91	5.57			

TABLE XVII. Values of MB_o min., R_v, and M. Solar altitude =90° (air mass=1.0), clear atmospheric conditions.

	Distant	Semidistant	Near	Close
	MB_o min.			
Unshaded	64	32	16	8
Light shade	32	16	8	4
Heavy shade	16	8	4	2
Dense shade	8	4	2	1
	$1/MB_o$ min., Relative—R_v			
Unshaded	1	2	4	8
Light shade	2	4	8	16
Heavy shade	4	8	16	32
Dense shade	8	16	32	64
	Scene index—M			
Unshaded	0	3	6	9
Light shade	3	6	9	12
Heavy shade	6	9	12	15
Dense shade	9	12	15	18

establishment of this correlation we are forced to depend largely upon the information published in a previous communication.[x] For convenience, the data applying to the present problem are shown in Table XV. The third column of this table gives the number of scenes used in obtaining the average value of observed B_o min. shown in the fourth column. It will be seen that the values of B_o min. for the semidistant and the near unshaded scenes are based upon relatively large groups of measurements and hence, presumably, are of somewhat greater reliability than the average values for the other scene structures.

The total number of scenes on which information of this characteristic is available is 124, all of which were photographed under clear atmospheric conditions, atmospheric index, A, equal to 0, and at latitude 40°N (approximately). These scenes were photographed at times distributed rather uniformly throughout all months of the year and at various hours of the day. The luminous density ratio, R_v, therefore, was in all cases somewhat greater than 1.0, corresponding to solar altitude less than 90 degrees, air mass greater than 1. Since we wish to base the final values of scene index, M, upon the condition under which luminous density arriving at the earth's surface is a maximum, it is necessary to apply a correction to the values of B_o min. shown in the fourth column of Table XV. This correction must be based on an estimate of the average luminous density ratio under which the 124 scenes represented by the data in Table IV were photographed. In Table XVI, which is a duplicate of the second section of Table X, are shown the luminous density ratios, R_v, for different months of the year and hours of the day at altitude 40°N (or 40°S). It is estimated that the 124 scenes, which provide our information concerning the relation between B_o min. and scene structure, were photographed at rather uniformly distributed times as represented by those values in Table XVI which are printed in boldface type. The average of these R_v values is 1.30. It is probable that the time distribution is not perfectly uniform and that the values for the hours of mid-morning and mid-afternoon should be given somewhat greater weight than those near noon and of early morning and late afternoon. We therefore shall make an arbitrary adjustment and use 1.25 as the most probable weighted mean of the R_v values applicable to the 124 scenes represented by the data in the fourth column of Table XV. By multiplying the observed values of B_o min. in the fourth column of Table XV by 1.25, those in the next column, MB_o min., are obtained. These specific values of minimum object luminance are those which, on the average, are associated with a solar altitude of 90 degrees. When the sun is at the zenith, with a clear atmospheric condition, the volume density of luminous energy *entering* the scene space is at its maximum value (see Table VII, Part I, p. 158) and is not dependent upon the various aspects of scene structure or of the directional lighting conditions. As the luminous energy penetrates into the scene space occupied by various groupings of terrestrial objects, its intensity and the directions of propagation are modified. The resultant values of minimum scene luminance are the maximal to be expected for the various aspects of scene structure. The term MB_o min. is defined as the *maximal value of minimum object luminance*, B_o min.

It will be seen that there is a very marked tendency for the values of MB_o min. to vary by consecutive powers of 2. For the purpose of the final evaluation of scene structure index, we consider it permissible to round off the values in the fifth column to those shown in the last column of the table.

In the first section of Table XVII, the numbers in boldface type are taken directly from the last column of Table XV. The remaining numbers shown in lightface type, needed to fill in the table, have been estimated from such information as is available. There is little doubt that the sequence of MB_o min. values for the various types of spatial structure (semidistant, near, etc.), when conditions of light, heavy, or dense shade prevail, should be the same as the sequence for the unshaded condition. This sequence, as mentioned previously, is represented by consecutive powers-of-2 steps. A similar reasoning leads us to believe that the sequence of values for the steps in the local illumination series should be the same, for the different types of spatial structures, as the observed values shown for the near type which indicate strongly consecutive powers-of-2 steps.

[x] See reference 10, p. 673, Table III, and p. 671.

In the second section of Table XVII are given the *relative* values of the reciprocal of MB_0 min. These are designated by the symbol R_s. The third section gives the final values of *scene index*, M, which is defined by the equation

$$M = 10 \log R_s. \tag{30}$$

VIII. DIRECTIONAL LIGHTING

Under circumstances where a major portion of the light incident on a scene comes from a source which subtends a small solid angle, it is obvious that B_0 min., as observed from the camera position, will depend to some extent upon the angle between the direction of this component of illumination and the line of sight. This occurs when the scene is illuminated by the unobscured or partially obscured sun and, to a lesser extent, when the sun is obscured, but there is a relatively small area of the sky which is much brighter than the other portions. In general, the directional index need be applied only to the atmospheric conditions designated as *clear*,[y] *light haze*, and *medium haze*.

A. Front Lighting

It is evident that if a scene is illuminated entirely by a small intense light source located *exactly* at the camera position, there will be no shadows visible to an observer also at the camera position. This is the condition of perfect *front lighting*. The luminance differences observed under these conditions are due to differences in the reflectance of the scene elements and to differences in the distance between the light source and the scene elements. This hypothetical condition is never encountered in practice.

It is customary to apply the term *front lighting* to many conditions which depart widely from those described in the previous paragraph. Let the angle between the line of sight, SO (Fig. 31), and the direction from which the approximately unidirectional illumination, indicated by the line PO, enters the scene space, be designated as Δ. As long as Δ does not exceed some limit, which can be specified only approximately and is in fact dependent somewhat upon the solar altitude, the condition is described as *front lighting*. The angle, Δ, is, of course, measured in the plane which contains the line of sight, SO, and the line, PO. In terrestrial photography, it is assumed that the line of sight is approximately parallel to the horizontal plane. When the sun is on the horizon the angle, Δ, lies in the horizontal plane, and when the sun is on the meridian it lies in a plane perpendicular to the horizontal plane. For intermediate solar positions, the angle, Δ, lies in a plane of which the inclination to the horizontal plane varies from 0 to 90 degrees. The angle, Δ, can, in general, be resolved into a horizontal component, $h\Delta$, obtained by projecting the angle, Δ, onto the horizontal plane and into a vertical component, $v\Delta$, which is obtained by projecting the angle, Δ, into a vertical plane, as illustrated in Fig. 31. It is apparent that the magnitude of the departure from perfect front lighting is dependent upon the angle, Δ, and therefore cannot be specified in terms of $h\Delta$ only. Thus, with $h\Delta$ remaining 0, Δ may increase to 90 or even 180 degrees, and it is obvious that the condition of front lighting changes into very oblique lighting and finally into back lighting with the attendant alteration in the magnitude of B_0 min.

A scene is said to be front lighted when $h\Delta$ is less than about 60 degrees, provided Δ is less than approximately 75 degrees. For relatively low solar altitudes, the difference between $h\Delta$ and Δ is small, and as long as each is less than approximately 60 degrees the condition of front lighting can be assumed. If $h\Delta$ remains constant at 60 degrees and the solar altitude, $v\Delta$, is allowed to increase, Δ must also increase. When Δ has reached a value of about 75 degrees, a further increase in solar altitude without a decrease in $h\Delta$ would produce an oblique lighting condition which no longer can be classified as front lighting. However, by allowing $h\Delta$ to decrease somewhat below the value of 60 degrees, a further increase in solar altitude, $v\Delta$, can be

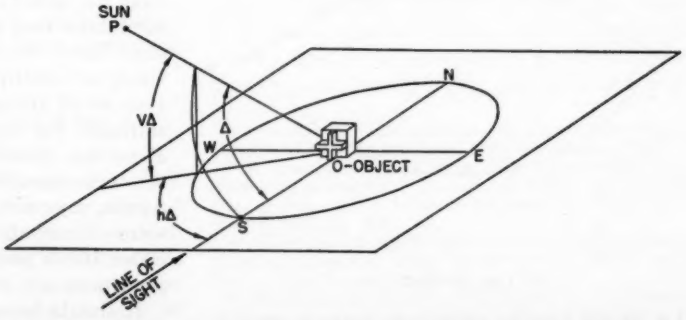


FIG. 31. Diagram illustrating conditions which control the directional viewing aspect.

[y] This includes a non-uniform cloud condition when the sun is unobscured and the sky in its immediate vicinity is clear.

tolerated without increasing Δ to a value greater than 75 degrees, and it appears that this condition may still be classified as front lighting, not involving a decrease in the value of B_o min. sufficient to call for an increase in the magnitude of camera exposure. However, if $h\Delta$ is decreased to zero, the solar altitude cannot be increased very much beyond 75 degrees without creating a condition of oblique lighting which does necessitate an increase in camera exposure. This particular condition of oblique lighting, that is, with the solar altitude somewhat greater than 75 degrees, gives rise to the condition which is frequently referred to as *top lighting*.

B. Side and Top Lighting

With relatively small values of solar altitude, $v\Delta$, a value of $h\Delta$ greater than 60 degrees leads to a condition of oblique lighting which is perhaps more generally referred to as cross lighting or side lighting. Rather long cast shadows extend laterally across the scene either to the right or left, with the consequence that B_o min., compared with that for front lighting condition, is decreased. This general condition persists with increasing solar altitude until the sun begins to approach the zenith. As this occurs, the cast shadows, instead of stretching horizontally or with small inclination to the horizontal direction, tend to point more and more downward toward the horizontal surface. When the solar altitude has exceeded some value, which, for all practical purposes, may be set at 75 degrees, the condition known as *top lighting* is encountered.

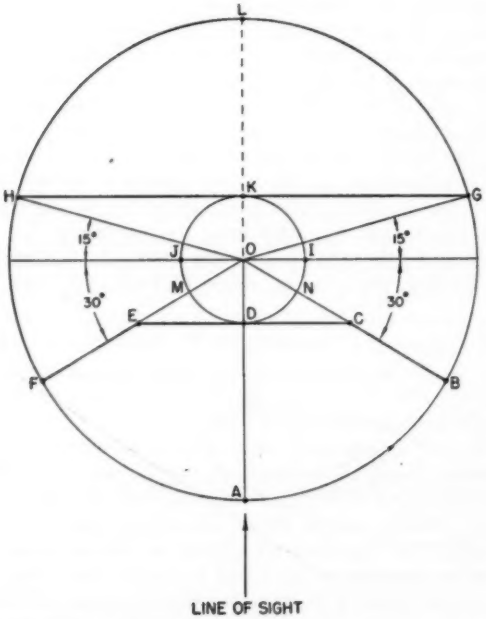


FIG. 32. Solid angular specifications for front, cross, top, and back lighting.

The effects upon B_o min. produced by top lighting and cross lighting are essentially similar, but there seems to be some evidence that the decrease in B_o min. (which is attributed to the directional lighting effect) with increasing values of $v\Delta$ (solar altitude) is somewhat less rapid than with increasing values of $h\Delta$. This may be due to a possible asymmetrical structure of scenes with respect to the vertical direction. Whatever may be the reason, it appears unnecessary to increase camera exposure because of high solar altitudes before a vertical component of Δ greater than approximately 75 degrees is encountered. The term *top lighting* is usually applied to this condition. From purely practical observations it seems to be well established that exposures in tropical countries need to be increased during noon-day hours somewhat above those used in regions of greater latitudes. It may therefore be concluded that frequent occurrence of very high solar altitudes leads to top-lighting conditions in tropical countries and this may account for the observed necessity of increasing somewhat the camera exposure in these latitudes.

C. Back Lighting

When the angle, Δ, has increased to a value somewhat greater than 90 degrees, the condition known as *back lighting* is encountered. The amount of singly or multiply reflected sunlight which reaches the shadow region in which the scene elements of minimum luminance are found is appreciably less than in the case of *side* or *top* lighting. This follows because those scene elements which receive direct sunlight are, in general, farther away from the scene elements of minimum luminance and hence subtend smaller solid angles at the shadow region. Moreover, with back lighting, the direction from which direct sunlight enters the scene space is such that the reflected sunlight tends to travel away from rather than into the shadow areas, as seen from the camera position. The elements of minimum luminance may receive some light directly from the sky. They are also illuminated in most cases by singly or multiply reflected skylight enhanced by very small amounts of greatly attenuated reflected sunlight. For back lighting it is necessary to give about four times the camera exposure required with the corresponding situation when front-lighted. Again, it is difficult to fix a precise value, but it seems reasonable to say that if Δ is greater than about 105 or possibly 110 degrees, the back-lighting conditions are realized.

It should be emphasized again that the angular values mentioned as differentiating front, cross, and

back lighting, can be considered only as approximate. There is no sharp line of demarcation between these conditions.

In Fig. 32 the approximate boundaries between these directional aspects of illumination are shown. This is a projection on the horizontal plane of the lines, on the upper hemispheric surface, which mark the boundaries of the solid angles embracing the types of directional illumination which have been discussed.

Thus, directional illumination entering the scene space along any line through O within the solid angle, the boundary of which is indicated by the lines, *FEDCB*, gives rise to the condition of front lighting.

The circle, *DJKI*, marks the boundary of the solid angle corresponding to top lighting. Thus, if the zenith distance of the sun does not exceed about 15 degrees, top lighting is encountered. The solid angle lying between the lines, *HKG* and *FEDCB*, exclusive of the circle, defines cross or side lighting. The solid angle included within the figure, *HKGLH*, defines the condition of back lighting.

The relative values of B_o min. for the three conditions (top lighting and cross lighting are considered identical insofar as B_o min. is concerned) must be considered as averages for the three solid angular specifications set up for their definition. Thus, if for a given scene structure, point of view, and atmospheric condition, the average value of B_o min. for all possible directions of the incidence of sunlight, within the angular limits for front lighting, is assumed to be unity, then the average B_o min. for the angular specifications of cross lighting is approximately $\frac{1}{2}$, while the average B_o min. for the conditions of back lighting is approximately $\frac{1}{4}$. Following the same method which has been used in expressing the light, atmospheric, and scene indices, the directional index, N, is obtained by the following relationship,

$$N = 10 \log R_d, \tag{31}$$

where directional lighting ratio, R_d, equals the reciprocal of the relative B_o min. value.

The values of relative B_o min., R_d, and N corresponding to the four directional lighting conditions are given in Table XVIII.

TABLE XVIII. Directional index, $N = 10 \log R_d$. Directional ratio, $R_d = 1/\text{relative } B_o \text{ min.}$

Directional aspect	Rel. B_o min.	R_d	N
Front lighting	1.0	1	0
Side lighting	0.5	2	3
Top lighting	0.5	2	3
Back lighting	0.25	4	6

IX. PHOTOGRAPHIC EFFICIENCY OF RADIANT ENERGY

There is one more subject that must be discussed before we can proceed with the computation of correct camera exposure by use of the data which have been presented in the preceding sections. This is the relation of the *photographic efficiency* of the *radiant energy* coming from the scene element of minimum luminance to that of the radiant energy which is used in the measurements of the *speed* of photographic negative materials.

The spectral composition of the radiant energy for use in the evaluation of the speed of negative materials was established by the action of the Seventh International Congress of Photography,[54] held in London July 9–14, 1928. The resolution establishing the photographic unit of intensity reads in part as follows: "The photographic unit of intensity for the sensitometry of negative materials shall be defined as the intensity of a filtered source of radiation having a luminous intensity of one international candle, and produced by a gray body at a color temperature of 2360°K, together with a selectively absorbing filter. . . ." The selectively absorbing filter referred to in this quotation is a two-solution liquid filter based upon the work of Davis and Gibson.[55]

This photographic unit of intensity has more recently been affirmed by the action of the American Standards Association[56] and by the British Standards Institution.[57]

Since the adoption in 1928 of the photographic unit of intensity mentioned above, two things have happened in the field of photometry which have some bearing on the precise specification of this unit. The international unit of luminous intensity, the international candle, is now defined in terms of a blackbody at the temperature of freezing platinum. This replaces the older unit which was maintained on a somewhat arbitrary basis by international agreement between national standardizing laboratories. This older unit was maintained in terms of standardized incandescent lamps. The adoption of the blackbody unit involves a relatively small, about 1 percent, change in luminous intensity. The other action which has a bearing on the photographic unit is the adoption of a new scale of color temperature. However, we interpret the action of

[54] *Proceedings VIIth International Congress of Photography*, London (W. Heffer and Sons, Ltd., Cambridge, 1929), p. 173.

[55] R. Davis and K. S. Gibson, "Artificial sunlight for photographic sensitometry," *Proceedings VIIth International Congress of Photography*, London (W. Heffer and Sons, Ltd., Cambridge, 1929), p. 161.

[56] "American standard method for determining photographic speed and exposure index," *American Standards Association Z38.2.1* (1947), approved July 15, 1947.

[57] British Standard 1380–1947. "Speed and exposure index of photographic negative material," British Standards Institution, London.

the Seventh International Congress of Photography as establishing the spectral composition of the radiant energy in terms of which the photographic unit is defined. This composition is that shown in curve *A* of Fig. 35. According to this interpretation, no change in the quality of the radiant energy is implied by the revision of the color-temperature scale. Some revision of the definition is necessary to make it consistent with the new color-temperature scale, which requires a tungsten lamp operating at a color temperature of, let us say, 2400°K (instead of 2360°K) to be used with the specified Davis-Gibson filter in order to achieve the required spectral composition.

As will be seen later (curve *A*, Fig. 35), the spectral composition of the radiant energy emitted by this source-filter combination approximates fairly closely that of direct sunlight (curve *B*, Fig. 35) transmitted through 1.46 air masses of clear atmosphere (see p. 135 in Part I for definition of the term, "clear atmosphere"). At latitude 40°, the annual average air mass through which direct noon sunlight travels to the earth's surface is 1.46, and hence it may be said that the spectral composition of the international photographic unit of intensity approximates quite closely that of the annual mean of noon sunlight at Washington, D. C., the latitude of Washington, D. C., being 38°53′N.

In evaluating the photographic efficiency of light sources, it is necessary to express "exposure" in terms of radiant energy units rather than in terms of light units, such as the meter-candle. We have defined *exposure* very specifically as the product of exposure time, t (seconds), and the illuminance, I (meter-candles), incident on the photographic material. Hence, for all photographic procedures which are designed to reproduce the visual appearance of objects, exposure is defined in terms of luminosity, that is, *light*. Light is a psychophysical concept and has been defined precisely, as follows: "Light is the aspect of radiant energy of which the human observer is aware through the visual sensations which arise from the stimulation of the retina of the eye."[5]

There are many wave-lengths of radiant energy which produce a latent image when they impinge on a photographic material but which do not stimulate the human visual system and hence are ineffective in contributing to the production of *light*. Hence, for such wave-lengths the luminous efficiency, K (in lumens per watt), is zero, while the photographic efficiency (in terms of photic response per watt) may have a high value. For the evaluation of photographic efficiency of various sources of radiant energy, it is customary to specify their emission characteristics in terms of spectroradiometric emission curves which show the relationship between radiant energy intensity, J, and wave-length, λ. This relationship may be expressed formally as

$$J=f(\lambda). \tag{32}$$

In expressing the "exposure" incident on a photographic material in terms of radiant energy, the unit usually employed is ergs per unit area. This situation gives rise to another problem in the establishment of an unambiguous terminology for photographic sensitometry. When a photosensitive material has been subjected to the action of those wave-lengths of radiant energy to which the retina of the human eye is not sensitive, it still seems reasonable to say that the photographic material has been *exposed*. But we have restricted the term *exposure*, E, by definition. Therefore, if illuminance for a given wave-length of radiant energy or for a wave-length band of heterogeneous energy is zero, the *exposure*, in terms of our definition, must also be regarded as zero. It seems wise to adhere to this restricted definition. It is obvious that there are at least two phases of exposure, considered as a general concept, with which we must deal in photographic sensitometry. The one which we have defined by the single word, *exposure*, might have been designated as *luminous exposure*. The two-word term, however, is less convenient and, in view of its predominant usage in the photographic sensitometry of all processes designed to reproduce the visual appearance of objects, we recommend that the single word, *exposure*, be used as already defined.

To meet the need for a term for use in referring to exposure, which must be expressed in terms of radiant energy units, we suggest *radiant exposure*. This can be defined in a fashion exactly parallel to the definition of *exposure*. The radiometric correlate of illuminance is irradiance, H (see Table I), which is expressible as ergs per second per square centimeter. The product of irradiance by exposure time, t, therefore becomes ergs per square centimeter, and we have the formal definition of the term:

$$\text{Radiant exposure}=H\cdot t(\text{ergs/cm}^2). \tag{33}$$

The *quality*, more specifically the *spectral composition*, of radiant energy reflected (or transmitted) by a scene element depends upon two factors: (1) the ability of the scene element to reflect (or transmit) the various wave-lengths of radiant energy incident thereon, and (2) the spectral composition of the incident radiant energy.

A. Spectral Composition of Radiant Energy Coming from the Scene Elements of Minimum Luminance

There seems to be no valid reason for assuming that the average (the statistical average of a vast

number of scenes covering the entire gamut of all possible structural variations) scene element of minimum luminance departs appreciably from the condition of non-selective reflection (or transmission). Within the spectral region with which we are concerned in this discussion (this region will subsequently be defined precisely), such surface elements do not reflect radiant energy of any particular wave-length more copiously than that of any other wave-length and hence, *on the average*, are non-selective reflectors and might appropriately be referred to as "photographically neutral" or "non-selective."

The *photographic efficiency*, W, of a source emitting heterogeneous radiant energy is dependent upon its *spectral composition* and upon the *spectral sensitivity* of the photographic material.

For the clear atmospheric condition, the spectral composition of sunlight arriving at the earth's surface for various solar altitudes is known with great certainty (see Fig. 5). Hence it is possible to compute the photographic efficiency, W, of sunlight, provided the *spectral sensitivity* of the photographic material is known. This characteristic of a photographic material is usually expressed graphically in the form of a curve showing the relationship between the sensitivity, C, for the various wave-lengths of radiant energy, as a function of wave-length,

$$C=f(\lambda). \tag{34}$$

The photographic efficiency, W, may then be expressed formally as

$$W=\frac{\int_0^\infty J_\lambda C_\lambda d\lambda}{\int_0^\infty J_\lambda d\lambda}. \tag{35}$$

When we attempt to evaluate the photographic efficiency of various phases of skylight, the situation is much less favorable because of the scarcity of comprehensive spectroradiometric data. While considerable work has been done on measuring the spectral composition of the radiant energy coming from the sky, the available data are not harmonious. There is disagreement between the results published by different observers, arising probably from the difficulty of specifying precisely the atmospheric conditions which existed when the measurements were made. These conditions are usually described qualitatively by such terms as "clear blue sky," "light haze," etc. Furthermore, in many cases data are presented without a definite statement of the exact location of the part of the sky on which measurements were made and the position of the sun at the time. On the whole, therefore, it is difficult, if not impossible, to use this available information for calculating the photographic efficiency of skylight as compared with that of the international photographic unit of intensity.

One of the best and most recent contributions to our knowledge concerning the spectral composition of various phases of skylight is that of Taylor and Kerr.[58] Unfortunately, as indicated by the title, their measurements do not cover the spectral region between 300 and 400 mμ, which is of great importance for the photographic problem with which we are concerned. However, we shall use some of their data for the illustration of certain points which will be discussed in the following pages.

We have already expressed the intention of assuming that, *on the average*, the scene elements of minimum luminance are non-selective reflectors or transmitters and therefore the spectral composition of the radiant energy reflected or transmitted by these elements is practically identical with that of the *incident* radiant energy. The scene elements of minimum luminance are seldom, if ever, illuminated by direct sunlight. In some cases they may be illuminated directly by radiant energy coming from relatively large solid angular areas of the sky, but in many cases by sky elements of relatively small angular dimensions. The radiant energy which is incident on these scene elements, therefore, may consist of: (a) sunlight which has been singly or multiply reflected from nearby objects which, *on the average*, are non-selective reflectors; (b) radiant energy coming directly from the sky; (c) skylight which has been singly or multiply reflected from nearby objects; or (d) mixtures in various proportions of any pair or of all three of the (a), (b), and (c) components. It appears, therefore, that, *on the average*, the radiant energy incident on the scene elements of minimum luminance is a very thorough mixture of that present in sunlight (the spectral composition of sunlight depending upon solar altitude) with that of the radiant energy coming from the entire sky hemisphere. The mixture proportion of the two components is probably approximately the same as the ratio between the volume densities of the radiant energies[a] arriving at the earth's surface from these two sources.

We have been unable to find in the literature any published data giving the spectral composition of this hypothetical mixture of sunlight and skylight. We can, however, from some of the available data draw some fairly reliable conclusions concerning the probable spectral composition of this radiant energy.

[58] A. H. Taylor and G. P. Kerr, "The distribution of energy in the visible spectrum of daylight," J. Opt. Soc. Am. **31**, 3 (1941).

[a] Radiant density, ergs per unit volume (the volume density of *radiant* energy), is the physical correlate of the psychophysical term *luminous density*, lumergs per unit volume (the volume density of *luminous* energy). See Table I.

First, let us consider the case of a completely overcast sky. As stated previously (p. 172 in Part I), the spatial distribution of the light transmitted by the cloud layer giving rise to this condition is illustrated at 12 in Fig. 22. Under these conditions, the maximum sky luminance is at the zenith and the position of the sun cannot be determined by visual observation since its location is not indicated by a sky region of enhanced luminance. The light traversing this cloud layer is almost completely diffused and the luminance of the sky decreases rather slowly from the zenith to the horizon.

Now, let us consider the composition of the radiant energy which is incident upon the upper boundary of the cloud zone corresponding to the "completely overcast" condition. First, there is direct sunlight which, if the sun is at the zenith, has been attenuated by its passage through 0.8 air mass. This attenuation is selective (i.e., a function of wave-length) since it arises almost entirely from molecular scattering in accordance with the Rayleigh fourth-power law. Thus, the shorter wave-lengths are scattered more copiously than the longer. The spectral composition of the radiant energy of the sunlight reaching the upper surface of the cloud zone (within the wave-length region with which we are concerned) for solar altitudes corresponding to greater air mass is shown by the curves 1, 2, 3, etc., in Fig. 5. It is evident from these that the relative proportion of radiant energies of the shorter wave-lengths, 300 to 400 mμ, is less than for those wave-lengths lying within the region between 400 and 700 mμ.

In addition to the direct sunlight there is incident upon the upper layer of the cloud zone radiant energy coming from all points of the entire sky hemisphere. This radiant energy consists of that which has been scattered out of the direct sunlight by molecular scattering in accordance with the Rayleigh fourth-power law. Compared with sunlight, this scattered radiant energy is much richer in those wave-lengths in the region between 300 and 500 mμ than in that between 500 and 700 mμ. If there were no deflected radiant energy having directions such that it passes outward into space, the mixture of this scattered radiation with the unscattered radiant energy in the direct sunlight should yield a spectral composition identical with that of the sunlight as it enters the earth's atmosphere (see curve *B*, Fig. 39). However, there is some back scatter and hence we cannot assume a spectral composition equivalent to that of the sunlight outside of the earth's atmosphere which, according to Moon,[3] is represented by the spectro-radiometric curve in Fig. 4.

Now it appears reasonable to assume that within the cloud zone there is no appreciable selective absorption of radiant energy (within the wave-length region with which we are concerned) nor is there any appreciable selective scattering other than the molecular scattering mentioned previously. The sizes of the water droplets which constitute the cloud mass are all appreciably greater than the particle sizes which result in the Rayleigh fourth-power scattering. Nor does it seem reasonable to expect that there is any *selective* reflectance of the radiant energy which enters the cloud zone. If these assumptions are correct and if we further assume complete diffusing and mixing of the two components within the cloud zone before they emerge and travel toward the earth, then we must conclude that the spectral composition of the radiant energy reaching the earth from the completely overcast sky is still not identical with the spectral composition of the radiant energy emitted by a blackbody at a temperature of 6500°K. Compared to a blackbody at 6500°K, the radiant energy from the overcast sky is relatively deficient in the regions of shorter wave-length, 350 to 450 mμ. The meager data available tend to confirm this conclusion. Thus, in Fig. 37, curve *C* is plotted from the measurements made by Taylor and Kerr on a completely overcast sky, while curve *A* is that for a blackbody at 6500°K. Unfortunately, the Taylor-Kerr data do not extend below wave-length 400 mμ, and we are of the opinion that there is some possibility that their values for the region 400 to 425 mμ may be subject to some uncertainties. The dotted section of curve *C*, which is designated as *C'*, 350 to 425 mμ, has been drawn in a manner which seems to us to be a probable representation of the facts. But the extrapolation over such a wide wave-length band is hazardous. However, we can check, at least approximately, the validity of this extrapolation by comparing the computed results (based on the curve, *CC'*) of relative photographic efficiency with those obtained by the direct sensitometric measurements which will be discussed shortly.

B. Sensitometric Measurement of the Relative Photographic Efficiency of Radiant Energy

Quite apart from the validity of our conclusions concerning the spectral composition of radiant energy reaching the earth from a completely overcast sky, we are of the opinion that this spectral composition, whatever it may be, represents closely the quality of the radiant energy which, on the average, is reflected from the scene elements of minimum brightness. This assumption is approximately valid for all values of solar altitude. As the air mass increases, the loss of radiant energy in sunlight is more rapid in the region between 300 and 500 mμ than in the region between 500 and 700 mμ. At the same time, however, the radiant

density resulting from direct sunlight decreases more rapidly than the radiant density due to skylight. Hence, a mixture of these two components (sunlight and skylight) for all solar altitudes tends to remain constant with respect to spectral composition.

A similar situation exists with respect to the constancy of the spectral composition of the radiant energy reflected from the scene elements of minimum brightness with various atmospheric conditions varying from clear to very cloudy. When the atmosphere is clear, the radiant energy incident on the shadow areas is made up of a mixture of skylight and of sunlight which has been singly or multiply reflected from nearby objects. The skylight component is relatively rich in the shorter wave-lengths as compared with the sunlight. As cloud masses begin to form, beginning with the light-haze condition, the skylight contains relatively less of the shorter wave-lengths, while the sunlight remains practically constant, for a given solar altitude, in spectral composition, but decreases rapidly in intensity. As the thickness of the cloud layer continues to increase, the same compensating effect persists. Finally, when the sky is completely overcast, the radiant energy reaching the earth from the sky is a thorough mixture of the sunlight and skylight incident upon the upper surface of the cloud zone. It is highly probable, therefore, that, irrespective of the atmospheric condition, the spectral composition of the radiant energy coming from the scene elements of minimum luminance approximates closely that of the completely overcast sky.

It must be kept in mind at all times that this communication deals with the determination of correct camera exposure for obtaining a negative from which a positive is to be made, i.e., a negative-positive process. Our data indicate conclusively that the minimum object luminance, B_o min., is the best criterion of the correct camera exposure. It seems rather unlikely, however, that B_o min. is the best criterion of correct camera exposure for a reversal process, either color or black-and-white. We are inclined to believe that, when the final evidence is available, the luminance of a scene element in the highlight region (not necessarily B_o max.) will be found to be the most reliable criterion of correct camera exposure. Even for a reversal process, the fundamental generalized criterion of correct camera exposure is still the *least* (modified by a suitable safety factor) which will yield a final result of excellent quality, the final result being the reversal positive.

Rather than rely on the results of computations by the usual mathematical methods of the relative photographic efficiency of this radiant energy, as compared with that of the international photographic unit of intensity, we have chosen to make some direct sensitometric measurements of the relative photographic efficiency of radiant energy coming from the zenith of a completely overcast sky, using several photographic materials differing in spectral sensitivity. For this purpose an intensity-scale sensitometer was used employing in one case the international photographic unit of intensity as a source, and in the other case radiant energy introduced into the sensitometer from the zenith of a completely overcast sky.

The sensitometric method of evaluating the photographic efficiency of an unknown source, x, of radiant energy relative to that of the photographic efficiency of the international photographic unit of intensity, y, is illustrated in Fig. 33. Curve Y is the usual D log$E$ characteristic obtained by using the reference standard, while curve X is the same function obtained by using radiant energy from a source of unknown efficiency. The points, a and b, are established on curve Y so that the gradient, dD/d log$E$, at the point, a, is equal to 0.3 of the average gradient of the curve Y between the points, a and b, this average gradient being equivalent to the slope of the straight line, ab. As indicated in Fig. 33, the points, a and b, are separated by a log$E$ interval of 1.50. The two points, m and n, on curve X are established in exactly the same way. The point, a, determines the value of exposure, E_a, used in computing the American Standard Speed for this material in accordance with the procedure adopted by the American Standards Association:[86]

$$\text{American Standard Speed} = 1/E_a. \qquad (36)$$

The position of the point, m, determines the value of exposure, E_m, and the *relative photographic efficiency*, RW_x, of the unknown source is given by the ratio of E_m to E_a,

$$RW_x = E_m/E_a. \qquad (37)$$

We have already mentioned the necessity of distinguishing carefully between radiant energy and luminous energy, that is, light, in discussing the photographic efficiency of light sources. This necessity arises because of the marked difference between the spectral sensitivity of the human eye and that of photographic materials. To illustrate this point and to establish the spectral regions with which we are concerned, Fig. 34 is given. The four curves, A, B, C, and D, show the relative spectral sensitivities of four general types of photographic materials which, for convenience, may be designated as high green-sensitive orthochromatic (curve A), normal or Type B panchromatic (curve B), high red-sensitive Type C panchromatic (curve C), and ordinary blue-sensitive materials (curve D). The ordinates, as shown at the left of the figure, are in relative terms and all four spectral-sensitivity

curves have been adjusted to have equal maximum ordinates. These curves are given simply for illustrative purposes and, while they were obtained by using individual commercial materials, they may also be considered to represent typical general classes. The spectral sensitivities of these photographic materials for wave-lengths less than approximately 400 mμ vary somewhat, but fall within the limits indicated by the dotted portions of the curves drawn to the left of the 410-mμ abscissa value, the solid line representing the average sensitivity in the region from 350 to 410 mμ. It will be seen that all photographic materials, irrespective of their sensitivity within the wave-length region from 450 to 700 mμ, have relatively high sensitivities in the region between 350 and 450 mμ.

Curve K is the standard luminosity curve for the average normal human eye. It has been plotted with its maximum ordinate equal to unity. This curve may be regarded as a representation of the functional relationship between retinal sensitivity, K (lumens per watt), and wave-length, λ, for the average normal human eye:

$$K=f(\lambda). \tag{38}$$

Curve T represents the spectral transmittance, T, of a typical photographic objective. This is expressed formally as

$$T=f(\lambda). \tag{39}$$

This has also been plotted with its maximum ordinate equal to unity. It is apparent from an examination of this curve that almost all radiant energy of wave-length less than about 340 mμ is absorbed by the type of glass lens with which the majority of cameras are equipped, and hence in evaluating the photographic efficiency of various light sources it is unnecessary to consider the spectral composition of the emitted radiant energy below wave-length of about 340 mμ.

On the other hand, the most red-sensitive of the photographic materials normally used for the reproduction of the visual appearance of objects has no appreciable sensitivity to wave-lengths of radiant energy greater than 700 mμ. Hence, in evaluating visual and photographic efficiencies of light sources, in terms of the spectral sensitivity of various types of photographic materials, it is only necessary to integrate values between approximately 340 and 700 mμ.

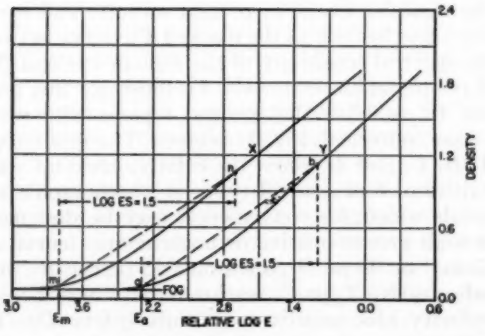


FIG. 33. D-log$E$ curves used in sensitometric measurement of the photographic efficiency of light sources.

Direct sensitometric measurements of the relative photographic efficiency of radiant energy from the zenith of a completely overcast sky were made in accordance with the method illustrated in Fig. 33. In making the sensitometric exposures for the overcast sky, a glass plate having a selective absorption approximately equivalent to that of a photographic objective (Fig. 34, curve T) was inserted between the source and the photographic material. The spectral sensitivities of the photographic materials used are those illustrated in Fig. 34 (curves A, B, C, and D). The values obtained are shown in the fifth column of Table XIX.

In current practice, the majority of negatives are made on materials of the types represented by Curves A, B, and C. The relative photographic efficiency values of these three materials are printed in boldface type and their average is shown at the bottom of the table. *We will use this average, 1.4, for incorporation in our final formula for the computation of correct camera exposure.* It will be seen that the radiant energy from the completely overcast sky, which, according to our assumption, is the same as that which is reflected, on the average, from the scene elements of minimum luminance, has a relative photographic efficiency about 40 percent greater than that of the international photographic unit of intensity.

C. Spectral Composition of Various Phases of Sunlight and Skylight

We shall now consider the spectral composition of the radiant energies reaching the earth in various phases of sunlight and skylight. In Fig. 35 the heavy solid curve, A, is the spectroradiometric emission curve for the international photographic unit of intensity. This, it will be remembered, is a combination of a graybody operating at a color temperature of 2360°K and a selectively absorbing filter. The dashed curve, B, represents the spectral composition of *annual mean* noon sunlight, with a clear atmosphere, at latitude 40°. An inspection of the two curves, A and B, along with the luminosity curve, K, in Fig. 34, indicates quite clearly that the spectral composition of the radiant energy emitted by the international photographic unit of intensity matches quite well that of mean noon sunlight in the wave-length region between 430 and

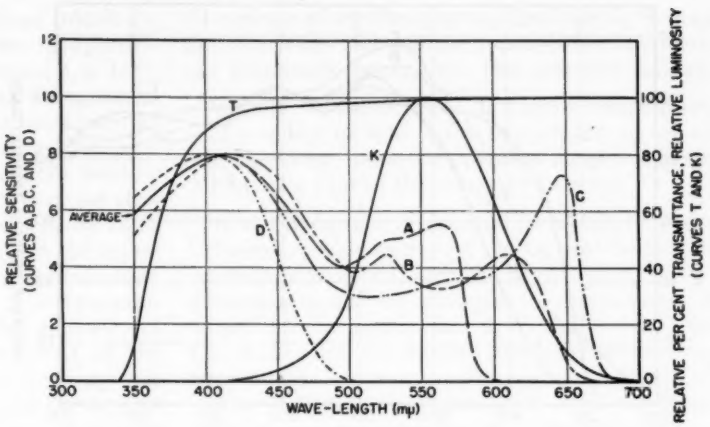


FIG. 34. Curves *A*, *B*, *C*, and *D* illustrate the relative spectral sensitivities of four typical photographic materials. Curve *K* is the spectral sensitivity of the average normal human eye (the luminosity function). Curve *T* is the relative spectral transmittance of a typical photographic objective.

650 mμ. In fact, direct visual comparison of these two sources shows that they are practically identical with respect to *color* (more specifically *chromaticity*), that is, visual appearance. Mean noon sunlight on clear days at latitude 40° has been assigned by various observers a *color temperature* of 5400°K. An inspection of curves *A* and *B* in the region of wave-lengths less than about 440 mμ, together with a consideration of the spectral sensitivities of the photographic materials shown in Fig. 34, leads to the conclusion that actual mean noon sunlight at latitude 40° should be somewhat more efficient photographically than the international photographic unit of intensity. Direct measurements have shown this to be the case although the difference is relatively small. Computation of the relative photographic efficiency, *RW*, of this phase of sunlight, using curve *B* (Fig. 35) and the spectral-sensitivity curves in Fig. 34, gives the values shown in the fourth column of Table XIX.

More important, however, for our purposes is a consideration of the spectral composition of other phases of sunlight and skylight. Curve *D* in Fig. 35 is derived from the measurements reported by Taylor and Kerr[58] for a mixture of sunlight and skylight on a horizontal plane, the measurements have been made on "clear" days. Curve *C* (Fig. 35) is also taken from the Taylor-Kerr data and represents the spectral composition of the radiant energy, evaluated on a horizontal plane, reaching the earth from a "completely overcast sky."

In general, the use of a *horizontal* plane in evaluating radiant energy or light for photographic purposes is subject to some objections. In the case of the completely overcast sky, however, the spectral composition of the radiant energy reflected from a non-selectively reflecting horizontal plane is very nearly identical with that of the radiant energy entering a unit volume of space at the earth's surface. In our opinion, the amount and spectral composition of the radiant energy entering a unit volume of space (i.e., radiant density) at the earth's surface is of maximum significance for the prediction of correct camera exposure, at least in the case of all negative-positive processes.

When the atmosphere is clear, the spectral composition of the radiant energy reflected from a non-selectively absorbing horizontal plane is not identical with that of the radiant energy entering a unit volume of space at the earth's surface. This statement applies to skylight or to mixtures of skylight and sunlight.

We have mentioned previously the desirability of knowing more precisely the spectral composition within the wave-length region from 340 to 440 mμ for the completely overcast sky condition. Having measured directly the relative photographic efficiency of this radiant energy, we are now in position to obtain a fair approximation to the desired information. By a process of trial and error we have established the extrapolated part of the curve (*C* in Figs. 35 and 37 and curve *J'* in Fig. 36) between 340 and 425 mμ which, by computation, gives approximately the same values of relative photographic efficiency as those obtained by direct measurements and shown in the fifth column of Table XIX.

For the sake of completeness, it may be well to

TABLE XIX. Photographic efficiency of light sources.

Photographic materials		Relative photographic efficiency—*RW*			
Spectral-sensitivity types	Curve	International unit	Mean Noon Sun, 40°	Overcast sky	*BB* 6500°K
Ordinary blue-sensitive	*D*	1.0	1.35	2.0	4.0
High green-sensitive orthochromatic	*A*	1.0	1.21	1.5	2.6
Normal sensitivity panchromatic (Pan Type *B*)	*B*	1.0	1.14	1.4	2.2
High red-sensitive panchromatic (Pan Type *C*)	*C*	1.0	1.10	1.3	2.2
Mean		1.0		1.4	

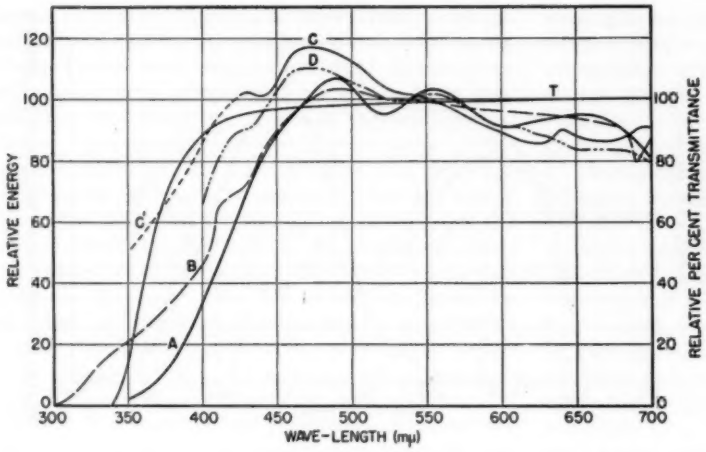


FIG. 35. Curve A, spectral composition of radiant energy emitted by the international photographic unit of intensity. Curve B, spectral composition of radiant energy in mean noon sunlight at latitude 40° (clear atmosphere). Curve CC', spectral composition of the energy from completely overcast sky; dotted portion extrapolated. Curve D, spectral composition of radiant energy reflected from a non-selectively absorbing horizontal plane illuminated by sunlight and skylight (clear atmosphere). Curve T, relative spectral transmittance of a typical photographic objective.

set forth briefly the operations involved in such computational work. In Fig. 36 the necessary functional relationships are shown. The significance of the various curves is indicated in the legend below the figure. The luminous efficiency of the international unit is given by:

$$K=\frac{\int_0^\infty J_\lambda K_\lambda T_\lambda d\lambda}{\int_0^\infty J_\lambda T_\lambda d\lambda}$$

$$=\text{luminous efficiency}^{aa}\ \text{(lumens/watt)}. \quad (40)$$

[aa] The luminous efficiencies, K and K', as defined by Eqs. (40) and (41), respectively, are the luminous efficiencies of the radiant energies actually incident on the negative material in the camera equipped with a lens (glass plus possible cementing material plus possible surface coatings for the reduction of reflection), the spectral transmittance of which is defined by the relationship

$$T=f(\lambda). \quad (39)$$

The luminous efficiency of the radiant energy *emitted* by the international unit is given by:

$$K_a=\frac{\int_0^\infty J_\lambda K_\lambda d\lambda}{\int_0^\infty J_\lambda d\lambda}. \quad (40a)$$

The luminous efficiency of the radiant energy *emitted* by the overcast sky is

$$K_a'=\frac{\int_0^\infty J_\lambda' K_\lambda d\lambda}{\int_0^\infty J_\lambda' d\lambda}. \quad (41a)$$

The differences between K and K_a and between K' and K_a', due to the insertion of the lens transmittance function as defined by Curve T in Fig. 36, are probably negligible. However, the T function actually limits the *effective* spectral sensitivity of photographic materials in the region between 300 and 400 mμ. Hence, it is necessary to introduce the T function into Eqs. (42) and (43) in order that the resultant relative efficiency, RW', of the radiant energy from the overcast sky shall be that which is actually realized in the exposure of a negative material in a camera. For the sake of uniformity, it seems desirable therefore to introduce the T function in all cases.

The luminous efficiency of the radiant energy coming from the overcast sky is given by

$$K'=\frac{\int_0^\infty J_\lambda' K_\lambda T_\lambda d\lambda}{\int_0^\infty J_\lambda' T_\lambda d\lambda}=\text{luminous efficiency}.^{aa} \quad (41)$$

If the two curves, J and J', are so plotted that $K=K'$, then the efficiency of the radiant energy coming from the overcast sky relative to that of the photographic unit is given by:

$$RW'=\frac{\int_0^\infty J_\lambda' C_\lambda T_\lambda d\lambda}{\int_0^\infty J_\lambda C_\lambda T_\lambda d\lambda}$$

$$=\text{relative photographic efficiency}. \quad (42)$$

However, if K is not equal to K', then

$$RW'=\frac{\int_0^\infty J_\lambda' C_\lambda T_\lambda d\lambda}{\int_0^\infty J_\lambda C_\lambda T_\lambda d\lambda}\cdot\frac{K}{K'}. \quad (43)$$

In Fig. 37 the Taylor-Kerr[58] overcast-sky data, as extrapolated in the manner just described, have been replotted as Curve C.

D. Color Temperature, T_c

Taylor and Kerr[58] report that the light from a completely overcast sky has a *color temperature* of

6500°K. Other observers have reported color-temperature values for the overcast sky of approximately the same magnitude (see reference 2, p. 138). Curve *A*, Fig. 37, shows the spectral composition of the radiant energy emitted by a blackbody at 6500°K and this has been so adjusted with respect to curve *C* that it indicates the probable validity of 6500° as the color temperature for overcast skylight. However, it becomes quite evident by comparing the ordinates of curves *C* and *A* in the region between 340 and 440 mμ that the two sources must be expected to have very different relative photographic efficiencies. The photographic efficiency of a blackbody at 6500°K relative to that of the international photographic unit of intensity has been computed. The results are shown in the last column of Table XIX. It will be seen that these values are all much higher than those for the radiant energy coming from the completely overcast sky (column 5, Table XIX). This emphasizes the inadequacy of color temperature as a significant index of the photographic efficiency of radiant energy from the overcast sky or of its spectral composition.

In view of this situation, it seems appropriate now to direct attention to the term *color temperature* and to its possible significance as an index of spectral composition and of photographic efficiency. This term has been appearing more and more frequently during recent years in the literature of photography. The concept as defined and properly used may be useful for some purposes. It has, in many cases, been misused and such misuse has led to much confusion and to many erroneous conclusions.

Let us look then at the definition of this term which has been formulated by the Nomenclature Committee of the Illuminating Engineering Society and which has received the approval of the American Standards Association. The definition follows:

> "COLOR TEMPERATURE: The color temperature of a source of light is the temperature at which a blackbody must be operated to give a color matching that of the source in question."[59]

If the definition of *color* as formulated by the Colorimetry Committee of the Optical Society of America is accepted, then the word "color" in this definition should be changed to *chromaticity*. It would then be appropriate to change the designation of the color-temperature concept to *chromaticity temperature*.

The definition as formulated used the term *matching* without qualification. It thus implies that the chromaticity of light from the source in question should differ from that emitted by a blackbody at some temperature by less than a *just-noticeable* chromaticity difference. While this condition is not impossible to realize in practice, it is, in fact, of relatively rare occurrence.

An explanatory sentence appearing in fine print below the definition as given reads as follows:

> "Color temperatures are usually assignable only for sources which have a spectral distribution of energy not greatly different from that of a blackbody."

This interpretation of the definition contains a qualitative specification "not greatly." Some individuals who are very conscientious in the application of this requirement may tolerate only a very small chromaticity difference. Others less careful and perhaps some overanxious to obtain a single number which may be used as a specification of

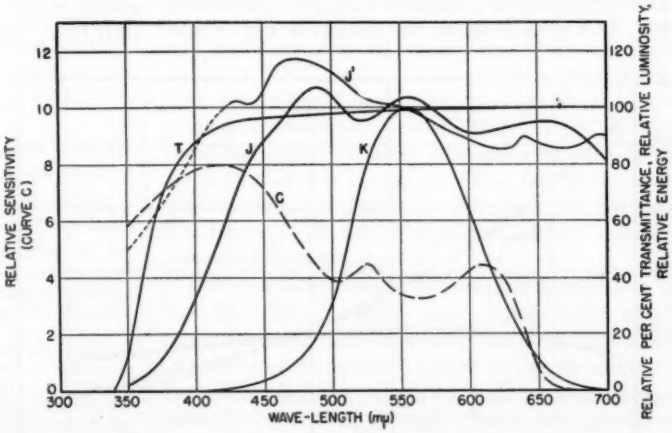


FIG. 36. Various functional relationships required for computation of photographic efficiency of radiant energy. Curves *C* and *K*, spectral sensitivity of a typical panchromatic material and of the average normal human eye, respectively. Curve *T*, relative spectral transmittance of a typical photographic objective. Curve *J*, spectral composition of radiant energy emitted by the international photographic unit of intensity. Curve *J'*, spectral composition of radiant energy from completely overcast sky.

[59] *American Standards Association Z7.1* (1942). Illuminating Engineering Nomenclature and Photometric Standards, approved February 27, 1942, by the American Standards Association as an American Standard. Prepared under the direction of and issued by the Illuminating Engineering Society, New York.

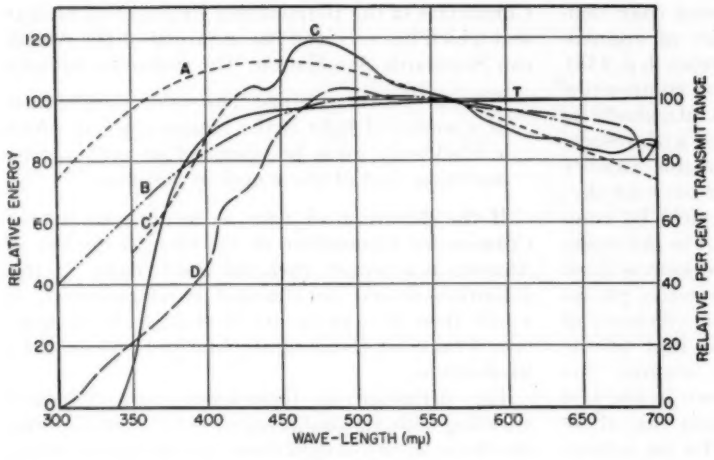


FIG. 37. Spectral composition of radiant energy from completely overcast sky, curve *CC'*, compared with that emitted by a blackbody at a temperature of 6500°K, curve *A*. Spectral composition of radiant energy in mean noon sunlight at latitude 40° (clear atmosphere), curve *D*, compared with that emitted by a blackbody at a temperature of 5400°K, curve *B*. Curve *T*, relative spectral transmittance of a typical photographic objective.

chromaticity, or even of spectral composition, may tolerate a relatively large chromaticity difference.

1. Various phases of sunlight and skylight.—In Fig. 37, in addition to curves *A'* and *C* already discussed, is shown the spectral composition of radiant energy of direct sunlight, curve *D*. This is the mean value at noon for latitude 40° (annual average air mass=1.46) for the clear atmospheric condition. Values of color temperature for this condition, as given by different observers, vary somewhat, but 5400°K appears to be a reasonably reliable average. Curve *B* shows the spectral composition of the radiant energy emitted by a blackbody at a temperature of 5400°K. The ordinates of the two curves, *B* and *D*, are so adjusted that the value of *K* (Eq. (40)) for curve *B* is equal to *K'* (Eq. (41)) for curve *D*. An inspection of curves *B* and *D* indicates the probable validity of 5400°K as the color temperature of this phase of sunlight.

Two other phases of natural light are represented in Fig. 38. Curve *B* shows the results of Taylor and Kerr's measurements on the radiant energy from a non-selectively reflecting plane surface inclined at an angle of 45 degrees relative to the horizontal plane and facing due north. The test plane was illuminated entirely by light from a clear sky (angular subtense of 1.5π-steradians), direct sunlight having been completely excluded. For this condition they report a color temperature of 12,600°K. The spectroradiometric curve for a blackbody at this temperature is shown as curve *A*.

Curve *D*, also from the Taylor-Kerr publication,[58] shows the spectral composition of the radiant energy from a non-selectively reflecting horizontal plane illuminated by direct sunlight plus skylight from "a fairly clear sky." These measurements were made near Cleveland, Ohio (latitude 43°N), between the hours of 9:30 A.M. and 3:30 P.M. during the months of August and September. To this mixture of sunlight and skylight Taylor and Kerr assign a color temperature of 6000°K. Curve *C* is that for a blackbody at a temperature of 6000°K.

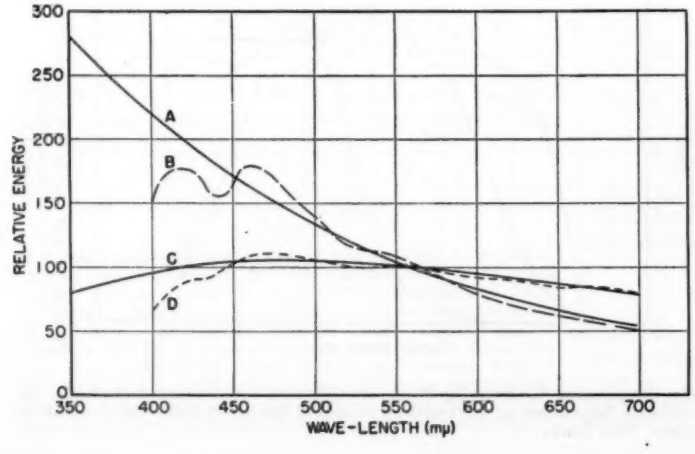


FIG. 38. Spectral composition of radiant energy reflected by a non-selectively absorbing plane surface facing north at an angle of 45° (clear atmosphere), curve *B*, compared with that emitted by a blackbody at a temperature of 12,600°K, curve *A*. Spectral composition of radiant energy reflected from a horizontal plane illuminated by direct sunlight and skylight from a "fairly clear" sky, curve *D*, compared with that emitted by a blackbody at a temperature of 6000°K, curve *C*.

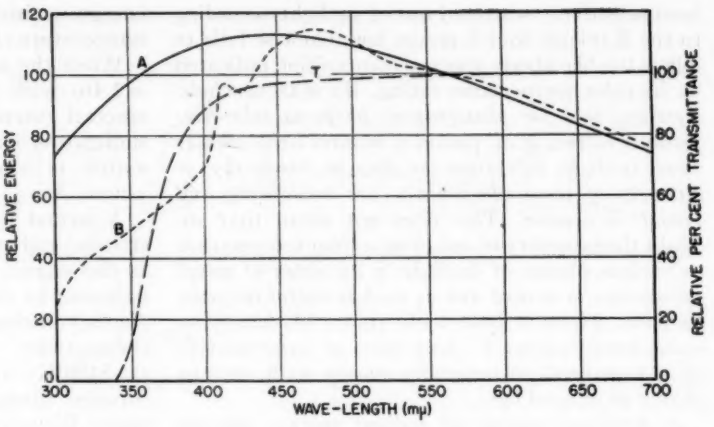


FIG. 39. Curve *B*, spectral composition of radiant energy in sunlight outside the earth's atmosphere. Curve *A*, spectral composition of radiant energy emitted by a blackbody at a temperature of 6500°K. Curve *T*, relative spectral transmittance of a typical photographic objective.

In Fig. 39 curve *B* represents the spectral composition of radiant energy for sunlight outside the earth's atmosphere (air mass = 0). This is taken from Moon's publication.[3] According to his computations, the color temperature of this phase of sunlight is 6200°K. Kimball[60] reports values of T_c for direct sunlight, with clear atmospheric conditions, "from 1700°K 15 minutes after sunrise to 5400°K at midday. Extrapolation to zero air mass gives 6500°K." The experimental observations on which Kimball based these statements were made by Priest[61] in Washington, D. C., during February of 1920 and June of 1922. Priest's measurements also indicate a probable value of T_c equal to approximately 5800°K for the zenith sun, air mass = 1.0, with clear atmospheric conditions.

From the evidence at hand we believe that the most probable value of T_c for sunlight outside the earth's atmosphere is approximately 6500°K. Curve *A*, Fig. 39, shows the spectral composition of the radiant energy emitted by a blackbody at a temperature of 6500°K. In all cases shown in Figs. 37, 38, and 39, the blackbody curve associated with a given phase of natural light is so plotted that the luminous efficiencies, as represented by Eqs. (40) and (41), are equal.

In all cases, the spectral composition of radiant energy emitted by the blackbody departs very much from that of the corresponding phase of sunlight and skylight and mixtures thereof, particularly in the region between 300 and 450 mμ. Hence, on the basis of the qualifying statement concerning the permissibility of assigned color temperatures to certain sources (see previous quotation, p. 113), no value of color temperature should be assigned to any of these phases of sunlight and skylight. This statement is valid only because of the omission of a limiting statement from the explanatory sentence which should read:

> "Color temperatures are usually assignable only for sources which, within the wave-length band extending approximately from 400 to 700 mμ, have a spectral distribution of energy not greatly different from that of a blackbody."

Color temperature is an index of the *visual appearance* of *light* and not of a *physical characteristic* of *radiant energy*, for instance, its *spectral composition*. There is little doubt that values of color temperature assigned to the various phases of sunlight and skylight, and their mixtures, are more or less precise indices of the chromaticities of these respective radiant energy compositions. They have very little other significance since, as we have seen in Table XIX, such values of color temperature have very limited significance as indices of relative photographic efficiencies of these sources as compared with each other or with the standard photographic unit of intensity. Nor are they useful specifications, at least for photographic purposes (see Figs. 37, 38, and 39), of the spectral composition of radiant energy.

The reason for this lack of significance lies in the fact that the sunlight, even outside the earth's atmosphere, departs markedly from the blackbody condition in the spectral region where photographic materials all have relatively high sensitivities. As sunlight penetrates through greater and greater air masses of the earth's atmosphere, this deficiency becomes even further exaggerated. Radiant energy from the clear sky which consists of that which has

[60] H. H. Kimball, "Records of total solar radiation intensity and their relation to daylight intensity," Monthly Weather Rev. 52, 474 (1924).

[61] I. G. Priest, "A new study of the leucoscope and its application to pyrometry. VI. Experiments on leucoscopic determination of the quality of sunlight and its diurnal variation," J. Opt. Soc. Am. 4, 480 (1920); "The colorimetry and photometry of daylight and incandescent illuminants by the method of rotatory dispersion," J. Opt. Soc. Am. and Rev. Sci. Inst. 7, 1175 (1923); "Preliminary data on the color of daylight at Washington," J. Opt. Soc. Am. and Rev. Sci. Inst. 7, 78 (1923).

been selectively scattered out of sunlight according to the Rayleigh fourth-power law, likewise fails to follow the blackbody spectral composition indicated by its color-temperature rating. *We must conclude, therefore, that for photographic purposes color-temperature ratings of all phases of natural light, such as direct sunlight, light from the clear or cloudy sky, or mixtures of these components, are misleading and should be avoided.* This does not mean that we doubt the propriety of assigning a color temperature to various phases of daylight *if an index of visual appearance is desired and as such is useful for some purposes.* There is little doubt that a blackbody at some temperature, T, does give an approximate (if not perfect) chromaticity match with various phases of natural light.

2. Artificial sources of radiant energy, selective radiators.—Since this communication, as indicated by its title, purports to deal only with the characteristics of the various phases of the radiant energy coming from the sun and sky, it may not be appropriate to discuss here the characteristics of artificial sources of radiant energy. In this particular section, however, we cannot resist the temptation to digress from our chief topic to call attention to the misleading implications of the color-temperature concept when applied to many artificial sources.

It is possible to obtain a perfect chromaticity match between the radiant energy emitted by a blackbody at some temperature and that from a source emitting radiant energy in two or three very narrow and widely separated spectral regions. Two such cases are illustrated in Fig. 40. A hypothetical source, X_1, is emitting radiant energy at three different wave-lengths, as indicated by the lines, A, B, and C. The relative amounts of radiant energy emitted in these three narrow bands are 36.2, 56.0, and 107.4, respectively. With this source a *perfect chromaticity* match is obtained with the radiant energy emitted by a blackbody operated at a temperature of 5400°K.

When the solar altitude is 43 degrees (air mass =1.46) with a clear atmospheric condition, the spectral composition of radiant energy in direct sunlight is such that a very good chromaticity match is also obtained with the hypothetical source, X_1.

A second hypothetical source, X_2 (Fig. 40), is emitting radiant energy in exactly the same manner as the source, X_1, but, in addition, is emitting as indicated by the line, D, at $\lambda = 390$ mμ. This source, X_2, is emitting radiant energy which matches in chromaticity: (a) the source, X_1, (b) a blackbody at 5400°K, and (c) direct sunlight which has traveled through 1.46 air masses of clear atmosphere. We have then four sources to each of which may be assigned (on the basis of a chromaticity match) the same color temperature, but which have vastly different spectral compositions. It is obvious that direct sunlight having a color temperature of 5400° will have a photographic efficiency differing appreciably from that of the equivalent blackbody (see Fig. 37). It is also obvious that the value of T_c which might be assigned to the hypothetical sources, X_1 and X_2, is quite useless and, in fact, misleading as an index of the spectral composition of the radiant energy emitted by them and of their relative photographic efficiencies.

The hypothetical sources shown in Fig. 40 are extreme cases. There are, however, some widely used commercial light sources which resemble in some degree the cases shown in Fig. 40. Prominent among these sources are the mercury vapor lamps of various types which emit radiant energy as spectral lines at various wave-lengths. Under certain operating conditions, this *line* spectrum may be associated with a more or less intense spectrally continuous emission of radiant energy. In some

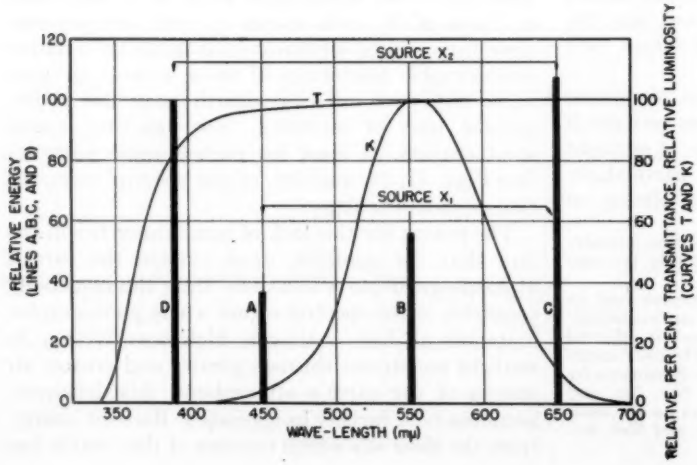


FIG. 40. Spectral composition of two hypothetical radiators of the line type. Light emitted by each of these sources, X_1 and X_2, produces perfect chromaticity matches with the light emitted by a blackbody at a temperature of 5400°K. Curve T, relative spectral transmittance of a typical photographic objective. Curve K, spectral sensitivity of the average normal human eye.

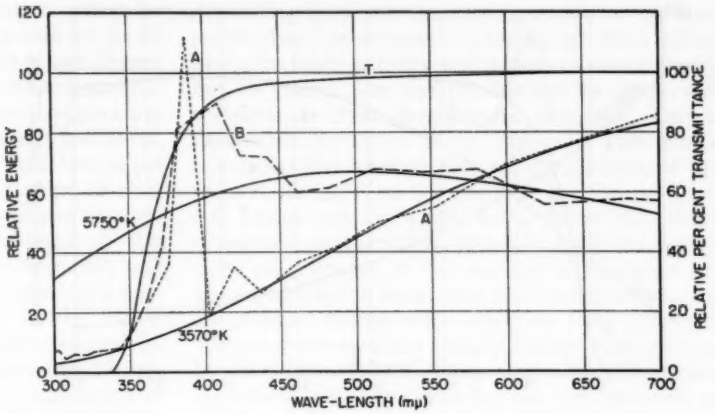


FIG. 41. Curve *A*, spectroradiometric curve for low-intensity carbon arc. Curve *B*, spectroradiometric curve for alternating-current high-intensity carbon arc (Strong Graf-Arc Lamp). Data furnished by courtesy of National Carbon Company. Curve *T*, relative spectral transmittance of a typical photographic objective.

cases, a fair approximation to chromaticity match may be obtained with a blackbody at some temperature, but such a color temperature is not a reliable index of relative photographic efficiency.

Another group of sources which have come into extensive commercial use during relatively recent years resemble to a lesser degree the line sources illustrated in Fig. 40, but which still are not sufficiently similar in their spectral-emission characteristics to a blackbody to warrant the assignment of color temperature as an index of photographic efficiency or of spectral composition. These are the fluorescent sources which are manufactured in a wide variety of types. There has been a tendency to ascribe a value of color temperature to some of these sources. It has been found, however, in almost all cases that these sources have relative photographic efficiencies very different from that of a blackbody at a temperature equal to the assigned color temperature.

The same situation exists also in the case of photoflash lamps in which various metallic foils and filaments of varying composition are burned in gas-filled enclosures. Here the spectrum is partially line and partially continuous, but the assignment of color temperatures to these sources is by no means a satisfactory index of their photographic efficiencies or of the spectral composition of the emitted radiant energy. We must also include in this same category those sources in which a high-potential, very short-time, electrical discharge is passed through a transparent envelope filled with vapors or gases or mixtures thereof.

The carbon arc is used extensively as an illuminant in various fields of photography, such as graphic arts, motion-picture studios, and for the projection of motion pictures in the theater. There are many forms of this source of radiant energy. For the purposes of this discussion, they may be classified roughly into three groups: (a) the low intensity arc, (b) the flame arc, and (c) the high intensity arcs, including both alternating current and direct current types.

The low intensity arc is no longer used very extensively in practical work. However, it is of sufficient interest from the point of view of this discussion to deserve our consideration. This arc uses electrodes which are composed of practically pure carbon and is, in most cases, operated on direct current. In this case, the major part of the light comes from the incandescent end of the positive electrode. In this region, the carbon electrode is heated by the passage of the electric current to a temperature approximating quite closely that at which carbon approaches the sublimation point. This emitted radiant energy has a spectral composition approximating quite closely that emitted by a blackbody at approximately 3570°K. It is therefore permissible to assign a color temperature to this particular part of the low intensity carbon arc emission. However, the arc stream between the ends of the electrodes emits a large amount of radiant energy in the region between 350 and 400 mμ.

We do not have at hand satisfactory data showing the relation between radiant energy intensity and wave-length for the *total* (incandescent electrodes plus arc stream) radiant energy emitted by a low intensity carbon arc. However, the general situation may be illustrated quite satisfactorily by curve *A* in Fig. 41. For the data used in plotting this curve, as well as those for curve *B* in Figs. 41 and 42, and curve *A* in Fig. 43, we are indebted to F. T. Bowditch of the National Carbon Company, and we herewith express our appreciation and thanks for his cooperation. In explaining the experimental technique used in obtaining these data, Bowditch states, "This curve represents the spectral energy distribution of a typical low intensity motion-picture projection arc, in which the positive crater

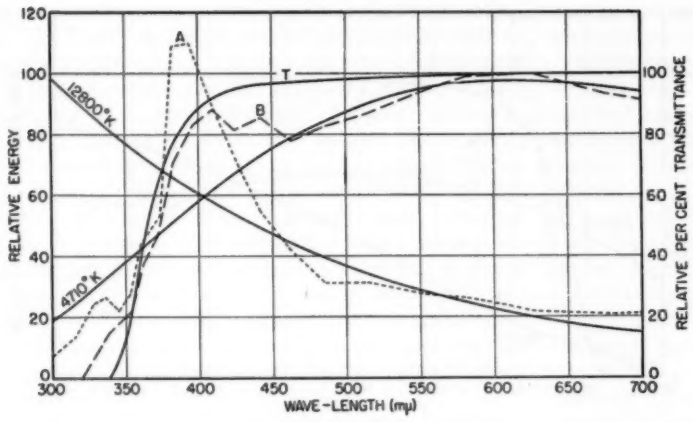


FIG. 42. Curve *A*, spectroradiometric curve for white flame arc (Sunshine Arc). Curve *B*, spectroradiometric curve for white-flame arc (motion-picture studio broadside unit). Data furnished by courtesy of National Carbon Company. Curve *T*, relative spectral transmittance of a typical photographic objective.

is placed at the focus of a mirror optical system. The measurements were taken, without this optical system, viewing the bare crater through a radiation shield which blocked out the tail flame from the view of the thermopile." The color temperature assigned to this source is 3570°K and the solid curve thus designated in Fig. 41 is that for a blackbody at that temperature. The position of this curve has been adjusted to the best fit with that of the energy emission from the low intensity arc in the visible region. It will be seen that the blackbody curve follows quite closely the radiant-energy curve for the arc between wave-length 400 and 700, and hence it seems entirely permissible to assign a color temperature to the *light* emitted by the low intensity arc. In other words, the chromaticity of the light emitted by the low intensity carbon arc matches quite satisfactorily that of the light emitted by a blackbody radiator at 3570°K. It will be seen, however, that there is also a rather large amount of radiant energy emitted by this low intensity carbon arc in the region between 350 and 400 mμ. This was found in spite of the fact that "a radiation shield which blocked out the tail flame from the view of the thermopile" was used. It seems reasonable to assume that if the experimental technique had been such as to measure *all* of the radiant energy coming from the arc stream in the region between 350 and 400 mμ, the intensity values would be appreciably higher than those shown in Fig. 41. Curve *T* in Fig. 41 represents the relative spectral transmittance of a typical photographic objective, and it will be seen that a very large proportion of the radiant energy emitted by the arc in the region between 350 and 400 mμ is transmitted by this lens and will therefore reach the negative material. It is obvious, therefore, that in the case of this low intensity carbon arc color temperature cannot be a satisfactory index of photographic efficiency or of the spectral composition of the total emitted radiant energy.

The flame arcs are obtained by using carbon electrodes which contain cores composed of various materials. Cores made by incorporating the rare earth metals of the cerium group produce a brilliant white flame which constitutes the chief source of emitted radiant energy. The incandescence of the carbon electrodes contributes a very small portion of the total emitted radiant energy. The emission characteristics of the white flame arcs may vary considerably, depending upon the amounts and proportions of the various flame materials which are incorporated in the core. In some cases (such as in the Sunshine arc), there is an extremely intense emission of radiant energy between 360 and 450 mμ, with very much less in the region between 450 and 700 mμ.[62] This is illustrated as curve *A* in Fig. 42. The color of the light emitted by this type of flame arc is "blue-white." It is not possible to obtain even an approximate chromaticity match between this source and a blackbody at any temperature, and hence color temperature is quite inadequate as a reliable index even of chromaticity as well as spectral composition and photographic efficiency. The blackbody curve in Fig. 42, designated as 12,800°K, is the closest approximation to a chromaticity match which can be obtained by the usual colorimetric methods of computing such color-temperature relationships.

Other types of white flame carbons which also use cores containing rare earth metals, such as cerium, emit radiant energy having a spectral composition quite similar to that of a blackbody at some temperature. Values of T_c varying from 4600 to 5500°K have been assigned to arcs of this type. These values in many cases are good indices of chromaticity and in some cases possibly fairly reliable as indices of photographic efficiency and spectral composition. A special case is illustrated in curve *B* of

[62] *Radiant Energy* (National Carbon Company Catalogue Section A-4000, 1944).

Fig. 42. The spectral-energy distribution is that for a complete arc unit consisting of a pair of flame arcs mounted in front of a fixed reflector and filtered by a diffusing glass placed between the arcs and the objects to be illuminated.[63] The color temperature assigned to the light emitted by this unit is 4710°K, and it will be noted that the curve thus designated in Fig. 42, which represents the emission characteristic of a blackbody at that temperature, fits very satisfactorily the radiant-energy emission of this unit in the wave-length region between 450 and 700 mμ. Even in this case, however, there is a considerable excess of radiant energy in the region between 350 and 450 mμ which is transmitted by the typical glass photographic objective, and hence it seems unlikely that the color temperature of 4710°K can be a very satisfactory index of the photographic efficiency or of the spectral composition of the radiant energy emitted by this flame arc unit.

The alternating current, high intensity arcs use carbons which are cored with materials similar to those used in the flame arcs, that is, combinations of the cerium group, but they are operated at very much higher current densities. A large part of the emitted radiant energy comes from the ends of the electrodes and originates largely from the gaseous material derived from the core elements rather than from the incandescent carbon itself. The spectral composition of the emitted radiant energy varies over a considerable range, depending upon the composition of the core materials and the operating conditions. In some cases, this spectral composition between 400 and 700 mμ is quite close to that of a blackbody at some temperature. Values of T_c can be assigned which vary from 5400 to 6400°K,[64] depending upon the exact composition of the core material and upon operating conditions. Such values of T_c are fairly satisfactory indices of chromaticity but, in general, are not reliable indices of spectral composition or of photographic efficiency. This is illustrated by curve *B* in Fig. 41. This curve represents the relative spectral distribution of the radiant energy from the arc with the Strong Graf-Arc Lamp and a reflector using 9-mm H.I. (high intensity) photographic carbons under the conditions as indicated in the caption of the figure. It should be kept in mind that this distribution is for the radiant energy after reflection from a mirror. The color temperature assigned to the light emitted by this unit is 5750°K. Here, again, it will be noted that the blackbody curve for this temperature, as shown in Fig. 41, matches the emitted radiant energy fairly well in the visible region, and hence it seems quite likely that the color temperature mentioned is a satisfactory index of the chromaticity of the light. However, here again, it will be noted that there is a large excess of energy, as compared with the blackbody at 5750°K, in the wave-length region between 370 and 440 mμ which makes it seem very unlikely that this color-temperature value can be considered as a satisfactory index of the photographic efficiency or of the spectral composition of the radiant energy emitted by this unit.

The direct current high intensity arc uses a positive carbon having a core containing the chemical combinations similar to those used in the flame-arc electrodes. The rare earth metals, such as cerium, lanthanum, neodymium, and praesodymium, have proved to be the best materials for incorporation in the cored carbons for arcs of this type. As is the

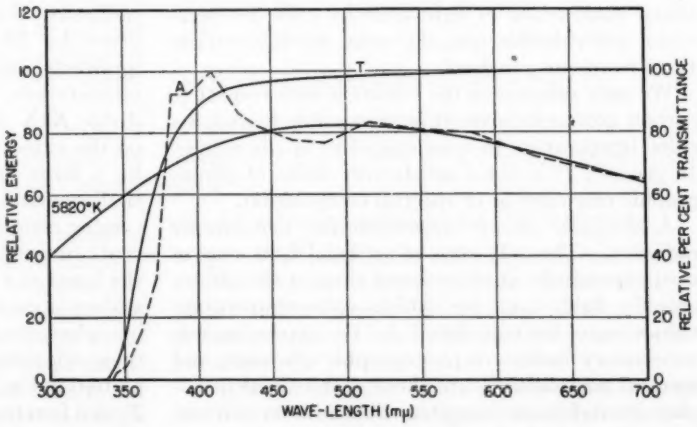


FIG. 43. Curve *A*, spectroradiometric curve for direct-current high-intensity motion-picture studio spotlight (Mole-Richardson Type 170 arc at full flood position). Data furnished by courtesy of National Carbon Company. Curve *T*, relative spectral transmittance of a typical photographic objective.

[63] F. T. Bowditch, M. R. Null, and R. J. Zavesky, "Carbon arcs for motion picture and television studio lighting," J. Soc. Mot. Pict. Eng. **46**, 441 (1946).

[64] F. T. Bowditch and A. C. Downes, "Spectral distribution and color temperatures of the radiant energy from carbon arcs used in the motion picture industry," J. Soc. Mot. Pic. Eng. **30**, 400 (1938).

case of the alternating current high intensity arc, the current density in the direct current high intensity arc is much higher than that found in the flame arcs. Curve *A* in Fig. 43 shows the spectral composition of the radiant energy emitted by an arc of this type to which a color temperature of 5820°K has been assigned. The spectroradiometric curve for a blackbody at this temperature fits quite closely with that of this particular direct current high intensity arc, and hence the color temperature is a fairly satisfactory index of the chromaticity of the light emitted by this source. Here, again, however, it will be noted that there is an excess of radiant energy emitted by this arc in the region between 380 and 440 mμ, compared with that emitted by the blackbody at a temperature of 5820°K. The color temperature which is an appropriate specification of the chromaticity of the *light* emitted by this arc cannot therefore be a very satisfactory index either of the spectral composition or of the photographic efficiency of this source.

We by no means wish to imply that the shortcomings of the color-temperature rating as a measure of spectral composition of the radiant energy emitted by carbon arcs have been overlooked by other workers. For instance, Bowditch and Downes,[84] in discussing this subject speak as follows: "Comparison of the eye and film response curves illustrates very well the very different results or effects of the radiant energy from a light source upon two photosensitive media whose spectral sensitivities differ considerably from one another. The futility of using color-temperature, which is a measure of the effect produced upon the eye, for the purpose of evaluating the effect of a light-source upon any other photosensitive material is thus apparent. . . . If one is interested only in the visual comparison of light sources, color-temperatures are valuable and do serve to differentiate them from one another."

We may summarize the situation with respect to carbon arcs as follows: It is permissible to assign a color temperature in *some* cases but not in others. In general, T_c is not a satisfactory index of photographic efficiency or of spectral composition.

3. Artificial sources approximating non-selective radiators.—The only class of artificial light sources used extensively at the present time in the photographic field, and for which color-temperature values may be considered to be approximately satisfactory indices of photographic efficiency and spectral composition, are those sources which employ incandescent tungsten filaments in various forms. Even with these sources, the color temperature is not a *precise* criterion of the spectral composition of the emitted radiant energy. These lamps are not blackbody radiators, nor are they even graybody radiators, but are *selective radiators*. The spectral emissivity of tungsten is a function of wave-length, and this functional relationship varies with the absolute temperature at which it is operated.

For most photographic purposes, the information desired concerning the radiant energy emitted by a source is its spectral composition. Color temperature was introduced as an approximate indication of this characteristic and for these lamps it serves the purpose with a fair degree of certainty. However, since color temperature is inadequate as a sufficiently precise index of spectral composition for so many sources used for various photographic purposes, much confusion might be avoided if the term were abandoned completely. There will undoubtedly be many objections raised to such abandonment, probably on the ground that it is desirable to have a single number which is a reliable indication of visual appearance (chromaticity) and of spectral composition. We admit the desirability of having such a single number, but if that single number in many cases gives misleading information, is there sufficient justification for its continued usage? The attempt to continue the use of color temperature for a single class of artificial light sources, that is, incandescent tungsten, will inevitably encourage the application of the same term to other sources, for many of which it has little, if any, significance except as an index of appearance.

Those who object to abandoning the term as applied to incandescent tungsten sources may very properly inquire what less objectionable term might be used as a substitute. Perhaps the answer to this may be found by considering the relation between various terms which may be used in describing the characteristics of incandescent tungsten radiators. Some significant data on this subject given by Moon[2] show the relationship between luminous efficiency, E (lumens per watt), true temperature, T, and color temperature, T_c. In Table XIX (see reference 2, p. 138) information on the values of these three characteristics is given for a large number of different light sources. We shall confine our attention entirely to that concerning commercially available gas-filled, incandescent, tungsten lamps. The values quoted refer to the lamp as a whole, that is, to tungsten filaments, seldom if ever straight but coiled in various ways, in a glass envelope containing a gas filling. To show these relationships graphically, the data have been plotted in Fig. 44. The circles show *true temperature*, T, as a function of lumens per watt, while the black dots show the relationship between color temperature, T_c, and lumens per watt. Each relationship can be represented quite satisfactorily by a straight line, and these lines are parallel to each other

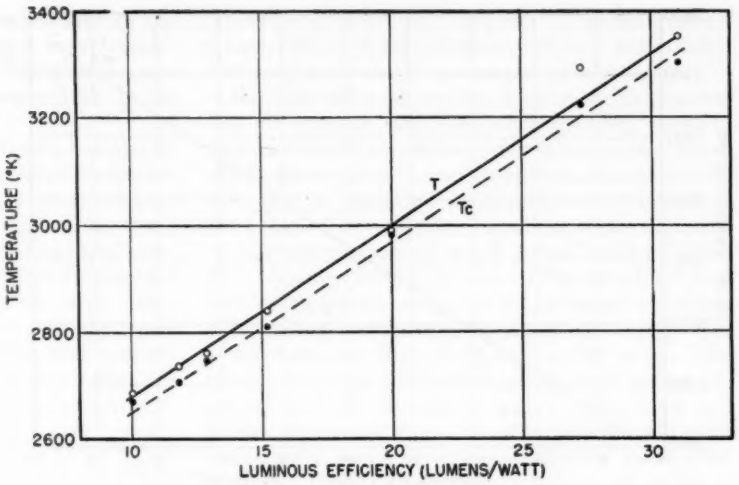


FIG. 44. Illustrates the approximately linear relationship between true temperature, T, and luminous efficiency, and between color temperature, T_c, and luminous efficiency for incandescent-tungsten lamps operating at various luminous efficiencies. Also shows the parallelism of the two linear relationships.

separated, on the average, by an ordinate value (temperature) of 34 degrees.

Under most favorable conditions in a high precision photometric laboratory, values of T_c may be measured to within ± 10°K (possibly ± 5°K). Usually, however, values of T_c are subject to uncertainties of about ± 20 or possibly ± 30°K. Hence, it appears that a value of *true temperature*, T, would be, for all practical photographic purposes, fully as satisfactory for the specification of the spectral composition of the radiant energy emitted by a tungsten lamp, as color temperature, T_c.

In our opinion, the term *temperature* would be somewhat less objectionable than the term *color temperature*, because *temperature* is descriptive of a purely physical characteristic of the radiator and is not a psychophysical concept, as is the case of color temperature. It is quite true that the same term, that is, true temperature, cannot be applied to many of the other sources which, as noted in previous paragraphs, are misrepresented by a color-temperature designation.

There is another possibility, namely, that of using the term *luminous efficiency* as a designation of the spectral composition of the radiant energy emitted by a tungsten lamp.

Forsythe and Adams[65] have published values of color temperature, T_c, and of luminous efficiency, lumens per watt, for a wide variety of incandescent-tungsten lamps. In Fig. 45 we have shown the relationship between luminous efficiency and T_c for those tungsten lamps which are used most extensively for the exposure of negative materials in the photographic field. We have not included all of the lamps on which Forsythe and Adams have given data because many of those are seldom, if ever, used in the making of photographic negatives in the camera. We have included all the lamps listed in Table IX (see reference 65, p. 61) under the heading "Some Characteristics of a Number of Lamps for Photographic Purposes." In addition to these, we have included all the general-purpose lamps (Table VIII, reference 65, p. 60) having wattage ratings of 100 and greater. We have also included those lamps in Table VIII designated specifically for use for flood lighting.

For these sources of radiant energy, the relation between luminous efficiency and color temperature is represented quite satisfactorily by a straight line. This substantiates the data previously shown in Fig. 44. According to the relationship shown in Fig. 45, color temperature is related to luminous efficiency by the following equation:

$$T_c = \text{luminous efficiency} \times 35 + 2286.$$

On the basis of the data shown in Fig. 44,

$$T = T_c + 34.$$

Hence, for the sources shown in Fig. 45, we may write

$$T = \text{luminous efficiency} \times 35 + 2286 + 34.$$

It appears that luminous efficiency is approximately as adequate an index of the spectral composition of the radiant energy emitted by these tungsten sources as either true temperature or color temperature.

Luminous efficiency is a psychophysical concept but it is one which can be applied to all light sources. It indicates specifically the *amount of light* which a source will produce for a given expenditure of energy, but it makes no invidious suggestion concerning the spectral composition of the radiant

[65] W. E. Forsythe and E. Q. Adams, "Color-temperature scale. Part II. Color temperature and characteristics of various popular lamps," Gen. Elec. Rev. **47**, No. 10, 59 (1944).

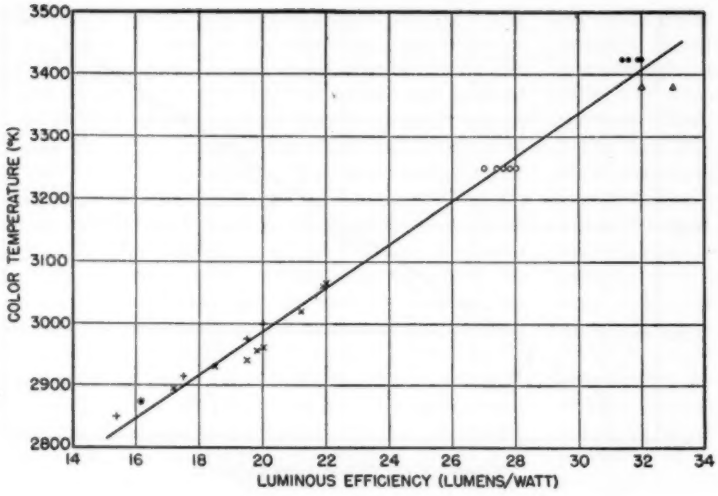


FIG. 45. Illustrates the rather close linear relationship between color temperature and luminous efficiency for incandescent-tungsten filament lamps used extensively in the photographic field:

× = General-service lamps.
+ = Flood-lighting lamps.
● = Photoflood lamps.
○ = Color photography lamps, 3250°K.
△ = Color photography lamps, 3380°K.

energy which is emitted by the source. Luminous efficiency is a reliable index of one of the three attributes of the color of light, namely, luminance, but gives no indication whatever of the other two attributes of the color of light, namely, dominant wave-length and purity. In our opinion, such a specification is more desirable than color temperature, which purports to give very much more information concerning the quality of emitted radiant energy than it actually does.

As a possible substitute for the term color temperature, *distribution temperature* has been suggested. This has some possible merit since it carries no implication concerning the appearance of the light emitted by a source. Such a term, however, could only be used as descriptive of the spectral composition of the radiant energy emitted by a source very similar to a blackbody with respect to emission characteristics. It would not be applicable to the various phases of natural light nor to many of the artificial sources currently used, with the possible exception of incandescent-tungsten lamps.

To digress somewhat further from the main theme, we may inquire as to the usefulness of the color-temperature concept in the field of color photography. The consequences of relying upon T_c as a reliable index of the spectral composition of a given sample of radiant energy are even more devastating in the field of color photography than in the case of the black-and-white processes where our chief interest in the light sources lies in their relative photographic efficiencies which must be considered in the prediction of correct camera exposure. It is true, even in black-and-white work, that sources must have a spectral composition of a certain character if colored objects are to be rendered in the reproduction in their correct tonal (visual appearance) relation to each other. In the case of color processes, however, a precise knowledge of spectral composition of the radiant energy emitted by a source is of great importance in determining correct color balance in the final reproduction, and in this field values of color temperature are likely to be extremely misleading if applied to sources which differ appreciably from a blackbody in their emission characteristics.

The significance and limitations of the color-temperature concept have been discussed at some length in the report of the OSA Committee on Colorimetry.[66] We are very much in accord with some of Moon's comments on this subject (see reference 2, pp. 130 and 131–132), from which we quote a few pertinent passages:

> "Color temperature has been very useful in engineering work and is employed extensively at present. Tables XVIII and XIX give data on some common sources. One cannot deny, however, that the whole concept of color temperature is on a very insecure foundation. First, the preceding definition assumes tacitly that an exact match can be obtained. It is well known that such is rarely the case. Radiation from a Planckian radiator will match the radiation from another Planckian radiator or from a graybody at the same temperature, but that is of no help, since color temperature then becomes the same as true temperature. Radiation from tungsten lamps can be matched, approximately but not exactly, with radiation from a Planckian radiator. . . . With some types of radiation, the concept of color temperature does not apply, since there is no possibility of getting even the roughest ap-

[66] Committee on Colorimetry, "Physical concepts: radiant energy and its measurement," J. Opt. Soc. Am. **34**, 183 (1944).

proximation to a match. With other sources the approximation is very good. But a very bothersome question is introduced: *How good* an approximation allows a color temperature to be assigned to a given radiation?

". . . the concept of color temperature has played an important part in the development of *incandescent* sources. Apparently, it was introduced as a rough engineering specification at a time when precise spectroradiometric measurements were almost prohibitively difficult. It still serves as a convenient *rough* engineering specification. *However, one may expect it to decline gradually, to be replaced by the much more satisfactory spectroradiometric curve and by the colorimetric methods* of Chap. XIII." (In all cases the italics are ours.)

Our feelings are somewhat stronger than those of Moon concerning the discontinuance of the usage of the term "color temperature," particularly in the field of photography. We should like to recast the last sentence of the above quotation to read as follows:

"However, we hope its use in the field of photography will decline rapidly or cease abruptly and be replaced by the much more satisfactory spectroradiometric curve, supplemented, if necessary or desirable, by classical colorimetric methods for the specification of chromaticity."

If, for any reason, a *color*, specifically *chromaticity*, specification is needed, we suggest the use of the generally accepted x and y values for the tristimulus coefficients or, provided agreement can be reached concerning the achromatic point in the chromaticity diagram, the even more meaningful specification of chromaticity in terms of *dominant wave-length* and *colorimetric purity*.

X. SAFETY FACTORS

Earlier in this communication we have mentioned, with only partial explanation, the desirability of introducing certain safety factors into the formula to be used for the computation of correct camera exposure. It has been shown (see Fig. 17 and Table VI), for instance, that, assuming the *average* conditions previously specified, the use of a constant value of 4.0 for the flare factor in Eq. (20b) leads to a value of camera exposure which is *just sufficient* to yield a negative from which an excellent print can be made. It must be recognized, however, that if the actual flare factor operative in a *particular* camera-lens-scene combination is less than 4, the camera exposure derived from the use of 4 as flare factor in the computation, will lead to a camera exposure somewhat too small to yield a negative from which an excellent print can be made. It was also shown (see Table VI and text) that, by introducing into the equation a partial safety factor, $SF' = 2.0$, the possibility of such undesirable underexposure would be completely eliminated.

If all the other factors involved in the determination of camera exposure were precisely known, there would be no necessity for the introduction of any further safety factor. In general, this is not the case and to take care of possible uncertainties in the values of several such factors, it seems desirable to introduce a second small partial safety factor, SF''. There is little objection to the introduction of such an additional safety factor, at least in the case of black-and-white negative-positive processes. This is emphatically illustrated in Fig. 46 at A. This curve shows the relation between log camera exposure and relative print quality. For very low values of camera exposure, negatives are obtained which yield prints of very poor quality, even when they are printed on the most appropriate grade of photographic paper available. As the camera exposure is increased, the quality of the best print which can be made from each of the negatives thus obtained increases up to a point where the *first excellent*, *FE*, print is secured. With further increase in the camera exposure, little, if any further improvement in print quality is obtained, and this continues until camera exposure has been increased to such an extent that print quality begins to decline. This point is indicated on the curve as *LE*, signifying the *last excellent* print. As will be seen, the ratio of the camera exposure yielding the last excellent print is 64 times that required to give the first excellent print. The curve shown is based upon the use of a modern negative material of high quality; a scene having the average luminance scale, $BS_0 = 160$; and a lens-camera-scene combination resulting in a flare factor of 4.0. Therefore, the camera exposure yielding the first excellent print can be doubled without any danger whatever of

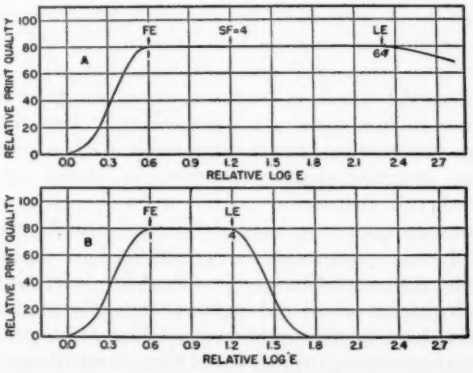


FIG. 46. Curve A shows the relationship between relative print quality and camera exposure for a black-and-white negative-positive process. Curve B shows the same relationship for a black-and-white reversal process.

loss of tone reproduction quality due to overexposure. Further increase in camera exposure may introduce certain other undesirable effects, such as negatives which are inconveniently dense and which perhaps may show enhanced graininess. We shall therefore adopt for the second partial safety factor, SF'', a value of 2.0.

While the camera exposure may be increased many times that required to give the first excellent print, without resulting in prints which show loss in tone reproduction quality, camera exposure much in excess of that demanded by the use of the two partial safety factors which we recommend should be avoided. Such large increase in camera exposure (provided that a lens aperture sufficiently small to give adequate depth of field is used) may demand an exposure time, t, so great that motion within the scene or movement of the camera may lead to blurring. On the other hand, if the exposure time is shortened so this does not occur, the f-value required may be so great as to entail undesirable decrease in the depth of field. Moreover, negatives which are unduly dense may require an inconveniently long time in printing, especially if projection printers are used and relatively great enlargements are desired. Negatives of excessively high density also increase the difficulties of retouching in case such hand work is necessary or desirable. In our opinion, the magnitudes of the safety factors which we recommend are not sufficient to result in serious trouble from the factors discussed in this paragraph.

While we are on this subject, it is desirable to call attention to the situation which may be encountered in black-and-white reversal processes where the introduction of safety factors of the magnitudes mentioned may be undesirable or inadmissible. The curve in Fig. 46 at B represents the quality curve for a typical black-and-white reversal process, and it will be seen that the last excellent print is obtained when the camera exposure is only four times as great as that resulting in the first excellent print. It is quite evident that, with this decreased latitude in permissible camera exposure, it may be inadvisable to apply safety factors of the same magnitude that are permissible in a black-and-white negative-positive process. We have introduced this discussion partly with the object of emphasizing the necessity of differentiating between *speed* and *exposure index* as generalized concepts.

The American Standard method for determining photographic speed and exposure index[56] previously referred to, applies only to "roll films, film packs, miniature camera films, sheet films, and plates intended for the making of monochromatic continuous-tone negatives in pictorial photography, exclusive of photography in the infrared." According to this standard, the value of American Standard Exposure Index is obtained by dividing the American Standard Speed by 4. This number, 4, is equal to the product of the two partial safety factors which we have indicated is our intention of using in the computation of correct camera exposure.

The introduction of these safety factors may appear to be a contradiction of our previous definition of the term, *correct camera exposure*, but, in fact, this is not the case. The definition as formulated previously (see p. 147 of Part I) is as follows: "Correct camera exposure is the one which is *necessary* and *sufficient* for the production of a negative from which an excellent print can be made. Any value of camera exposure less than this is insufficient and leads at once to loss of print quality. Any value greater than this is unnecessary." In our opinion, the introduction of reasonable safety factors in the formula for the computation of correct camera exposure does not invalidate the intent of our definition. Without the introduction of such safety factors, Eq. (20a) would give the correct camera exposure only in those cases where every variable which enters into the determination of camera exposure is known precisely and for those conditions of camera-lens-scene combination where the actual flare factor, FF, is 4.0 or greater. Actually, the accumulated uncertainties in these factors would probably lead to the same percentage of moderately underexposed negatives as of the moderately overexposed negatives in the sense of our definition of correct camera exposure. As stated previously, the adverse effects of somewhat greater camera exposure than the minimum required are of negligible importance. On the other hand, the effect of a similar degree of camera exposure less than the minimum required may lead to serious loss of tone reproduction quality. We take the position, therefore, that correct camera exposure should be defined in terms of an equation which includes a safety factor just sufficiently large to reduce the probable percentage of underexposed negatives to a very low value.

If we take the point of view that the speed of any photographic process should be expressed in terms of the least camera exposure which will result in a final reproduction of excellent quality, and if the processes differ among themselves with respect to the range of camera exposure giving results of excellent quality, it is obvious that the exposure index cannot in all cases be obtained by dividing the speed by a fixed value, for instance, 4.0. Therefore, while the ratio of speed to exposure index for a black-and-white negative-positive process may be 4, the ratio of speed to exposure index for a reversal process (either black-and-white

or color) may be less than 4 and, in fact, this is found to be the case. In general, then, speed and exposure index are related to each other by a safety factor, the magnitude of which must depend upon the nature of the process being considered.

XI. DETERMINATION OF CORRECT CAMERA EXPOSURE

Thus far our approach to the exposure problem has been analytical. We have discussed in detail all the factors which must enter into the determination of correct camera exposure and have given the numerical values for these factors which we consider to be sufficiently reliable for all practical purposes. These data, which are contained in the tables and graphs of the preceding sections, are voluminous and the reader may be perplexed with the seeming complexity of our treatment.

We shall now reverse the direction of approach and begin the integrating or synthesizing process involving simplification and the reduction of the information to a form sufficiently condensed for application to practice.

There are several methods by which the basic data given in the previous sections may be used for obtaining the desired value of correct camera exposure. We shall describe three such methods. The individual worker can then choose that method which, to him, seems most meaningful or intelligible and which is best suited to his mode of work and to his needs.

As a preliminary to the discussion of these methods, we shall review briefly the fundamental equations and the numerical constants which are common to all methods. First, let us look again at the equation which gives the relation between the illuminance, I_{io}, incident on the negative material in the camera and the luminance, B_o, of the corresponding scene element (see p. 128 of Part I):

$$I_{io}=B_o(F^2/4v^2f^2)\cdot\cos^4\theta\cdot H\cdot T_g\cdot 10.76; \qquad (3)$$

I_{io} is expressed in *meter-candles* and B_o is expressed in foot-lamberts. It should be remembered that I_{io} is the illuminance on the negative material due to light coming from the *object* and does not include the additional illuminance which may arise from non-image-forming *flare light*. By segregating those factors to which constant values may be assigned, Eq. (3) may be rewritten:

$$I_{io}=(B_o/f^2)\cdot K, \qquad (4)$$

where

$$K=(F^2/4v^2)\cdot\cos^4\theta\cdot H\cdot T_g\cdot 10.76. \qquad (5)$$

In a previous communication,[10] we concluded that a value of 1.25 for K should be satisfactory as representative of the average conditions which at that time existed in the field of amateur photography. Since the publication of that communication (1941), one change in the characteristics of photographic objectives has occurred which we believe should be recognized. We refer to the current practice of applying non-reflecting coatings to some or all of the free glass-air surfaces of the higher quality photographic objectives. These usually contain four or more coated surfaces so that the transmittance is appreciably increased as compared with a lens of the same structure having uncoated surfaces. A simple achromatic objective having only two free glass-air surfaces, even without coating, has a transmittance, T_g, of 0.90 or more. With current coating practice, the more complex objectives have, on the average, T_g values of from 0.80 to 0.90, depending somewhat upon the quality of the coating and the number of free glass-air surfaces. We previously assumed the most probable average value of T_g to be 0.70. We now believe that for current conditions the value of T_g should be raised to at least 0.85, thus making the new value of K equal to 1.52, which, for further use, we shall round off to 1.50. If T_g is assumed to be 0.90, K becomes 1.61.

In our previous evaluation of K, we took into account the falling off of illuminance at off-axis points and assumed a value of $\cos^4\theta=0.87$ and of $H=0.80$ (vignetting being due to the barrel of the lens mount) for a point 15 degrees off axis. There are some who do not favor the introduction of these values but prefer to evaluate K for object and image points lying on the optical axis of the photographic objective. However, we are still of the

TABLE XX. American Standard speeds and American Standard exposure indices.

Photographic materials	ASA speed S	ASA Exposure indices Z	LZ 10 Log$Z$
	40	10	10
	48	12	11
	64	16	12
Eastman Commercial Plate	80	20	13
	100	25	14
Kodak Panatomic-X Sheet Film	128	32	15
	160	40	16
Kodak Verichrome Roll Film and Film Pack	200	50	17
	250	64	18
	320	80	19
Kodak Super-XX Panchromatic Roll Film and Film Pack	400	100	20
	500	125	21
	640	160	22
Kodak Tri-X Panchromatic Sheet Film	800	200	23
	1000	250	24
	1250	320	25
	1600	400	26
	2000	500	27
		650	28

TABLE XXI. Values of MB_s min. Solar altitude = 90°.

Atmospheric condition	R_s	Scene structure: Distant Un.	Distant Lt.	Distant Hv.	Semidistant Un.	Semidistant Lt.	Semidistant Hv.	Near Un.	Near Lt.	Near Hv.	Close Un.	Close Lt.	Close Hv.
Clear	1	64	32	16	32	16	8	16	8	4	8	4	2
Light haze	1.4	46	23	11	23	11	6	11	6	3	6	3	1.5
Medium haze	2.0	32	16	8	16	8	4	8	4	2	4	2	1.0
Heavy haze / Light cloud	3.2	20	10	5	10	5	2.5	5	2.5	1.2	2.5	1.2	0.6
Medium cloud	5	12.8	6.4	3.2	6.4	3.2	1.6	3.2	1.6	0.8	1.6	0.8	0.4
Heavy cloud	10	6.4	3.2	1.6	3.2	1.6	0.8	1.6	0.8	0.4	0.8	0.4	0.2
Dense cloud	32	2.0	1.0	0.5	1.0	0.5	0.25	0.5	0.25	0.12	0.25	0.12	0.06

Note: For *dense shade*, divide value for *heavy shade* by 2.0.

opinion that for the amateur and commercial fields it is wise to evaluate K for an off-axis point which represents a reasonable average of the off-axis distance of all points in the entire negative.

In motion-picture studio practice where the lighting of the set is under precise control, and where a decrease in the illuminance incident on the negative material at off-axis points may be desired for aesthetic or compositional reasons, it may be preferable to evaluate K on the basis of axial transmittance. If this is done, θ becomes 0 and $\cos^4\theta = 1.0$. Also, for points on the axis there is no decrease in the illuminance incident on the negative caused by vignetting because of the lens barrel, and hence $H = 1.0$. Assuming a value of $T = 0.85$, the value of K becomes 2.18.

This paper is concerned chiefly with photography out of doors where the lighting is not under the control of the operator. Moreover, we have taken the position that the correct camera exposure is the least that will produce a negative from which an excellent print can be made. Thus, the use of the value of K based on the axial transmittance of the lens might, in some cases, lead to insufficient camera exposure at off-axis points. We shall therefore use the value of K based on the 15-degree off-axis transmittance of the lens, namely, $K = 1.50$.

We recognize the fact that in the photography of exterior scenes the amount of luminous energy available is frequently so great that it is common practice to close the iris diaphragm of a wide-aperture lens ($f/3.5$ for instance) to $f/8$ or even $f/16$. Under such conditions, the loss of illuminance, at off-axis points, caused by vignetting by the lens barrel is negligible, thus giving a value of $H = 1.0$ (rather than 0.8). In such cases, the incorporation of $H = 0.8$ in Eq. (5) will lead to somewhat greater camera exposures than are necessary. However, we take the position that it is preferable to follow a course which may give slightly denser negatives under conditions of high luminous density than to take the risk of giving too little camera exposure when the diaphragm must be set at its maximum aperture because the available luminous density is at a low level.

In Part I we have shown that it is necessary to introduce a *constant* value of flare factor, FF, in the equation for computing correct camera exposure and that the value of this factor should be 4.0. On page 152, we inadvertently implied that this choice of the constant value, $FF = 4.0$, for insertion in (20a) was made because 4.0 was the average value of flare factor which we had found to exist with the equipment at that time predominantly used in the amateur field. This implication, as a matter of fact, is misleading. The inherent necessity and the consequences of using the constant value of $FF = 4$ were analyzed in detail in the text descriptive of the graphic construction shown in Fig. 17 (Part I) and of the numerical values given

TABLE XXII. Values of $EC = f^2/t$.

Aperture ratio f	Exposure times, t, in (a) milliseconds (0.0001) and (b) fractions of 1 sec. (a) (b) 1 / 1/1000	2 / 1/500	4 / 1/250	8 / 1/125	16 / 1/64	32 / 1/32	64 / 1/16	125 / 1/8	250 / 1/4	500 / 1/2	1000 / 1
1	1,000	500	250	125	64	32	16	8	4	2	1
1.4	2,000	1,000	500	250	125	64	32	16	8	4	2
2	4,000	2,000	1,000	500	250	125	64	32	16	8	4
2.8	8,000	4,000	2,000	1,000	500	250	125	64	32	16	8
4	16,000	8,000	4,000	2,000	1,000	500	250	125	64	32	16
5.6	32,000	16,000	8,000	4,000	2,000	1,000	500	250	125	64	32
8	64,000	32,000	16,000	8,000	4,000	2,000	1,000	500	250	125	64
11	125,000	64,000	32,000	16,000	8,000	4,000	2,000	1000	500	250	125
16	250,000	125,000	64,000	32,000	16,000	8,000	4,000	2000	1000	500	250
22	500,000	250,000	125,000	64,000	32,000	16,000	8,000	4000	2000	1000	500
32	1,000,000	500,000	250,000	125,000	64,000	32,000	16,000	8000	4000	2000	1000

in Table VI. The fundamental reason for the use of $FF=4.0$ in Eq. (20a) is inherent in the conditions which determined the flare factor that existed with the camera-lens-scene combination used in making the negatives which led to the original establishment of the fractional-gradient criterion of effective camera speeds of negative materials. The validity of this criterion, $0.3\bar{G}$, was later proved by additional experimental work,[12] consisting in the making of a large number of psychophysical judgments of print quality using camera-lens-scene combinations involving values of flare factor other than 4.

The use of a value of $T_g=0.85$, instead of 0.70, will lead to a somewhat lower value of camera exposure, t/f^2. Furthermore, it is probable, as mentioned previously, that the present prevalent use of lenses having surfaces treated with low reflectance coatings may result in an average value of flare factor somewhat less than 4.0. This might lead to the impression that the value of flare factor incorporated in Eq. (20a) should consequently be reduced to a correspondingly low value. This conclusion, however, is quite erroneous since, as emphasized previously, the reason for using a value of FF equal to 4 in Eq. (20a) arises, not from any statistical average value of flare factor, but from the actual flare conditions which existed in the camera-lens-scene combination used in the establishment of the criterion on which the speed of negative materials is based. The probability of the more frequent occurrence of actual flare factor values of less than 4.0, because of the lower average surface reflectance in coated lenses, makes it all the more necessary to incorporate the partial safety factor of $SF'=2.0$ in the final equation for the computation of correct camera exposure.

Evidence has also been presented which indicates that the average photographic efficiency, RW (relative to the photographic efficiency of the radiant energy used in the measurement of the speed of negative materials), of the radiant energy coming from the scene elements of minimum luminance is 1.40.

Equation (20a) (p. 134 of Part I) gives the evaluation of camera exposure, CE, as

$$CE=\frac{t}{f^2}=\frac{1}{B_g\,\text{min.}\cdot FF\cdot K\cdot S}. \tag{20a}$$

This must now be modified by introducing RW into the denominator. Thus,

$$CE=\frac{t}{f^2}=\frac{1}{B_g\,\text{min.}\cdot FF\cdot K\cdot RW\cdot S}. \tag{20e}$$

By using the values $FF=4$, $K=1.50$, and $RW=1.40$, camera exposure reduces to

$$CE=\frac{t}{f^2}=\frac{1}{8.4B_g\,\text{min.}S}. \tag{20f}$$

Furthermore, it has been shown that it is desirable to introduce certain safety factors in order to avoid the possibility of underexposure in all cases where the values of the various factors entering into the determination of camera exposure may be subject to some uncertainty. The total safety factor, SF, which we recommend is the product of two partial safety factors, SF' and SF'', for each of which we recommend values of 2.0, thus making the *total safety factor* equal to 4.0. This number, of course, must be introduced in the numerator of the second member of Eq. (20f).

The information which is more generally available concerning the sensitivity of negative materials is *exposure index*, Z. This is derived from the speed by the relationship,

$$Z=S/4. \tag{44}$$

The substitution of Z for S in the denominator of Eq. (20f) is equivalent to the introduction of the safety factor, 4.0, in the numerator. Hence, the final and simplest form of the camera exposure equation becomes

$$CE=\frac{t}{f^2}=\frac{1}{8.4B_g\,\text{min.}Z}, \tag{20g}$$

which may be rewritten in the form

$$CE=\frac{1}{B_g\,\text{min.}}\cdot\frac{1}{8.4Z}. \tag{20g}$$

Values of exposure indices for various photographic materials are usually published by the manufacturers of these products. They can be measured by following the procedure set forth in detail in the Speed Standard as adopted by the American Standards Association.[56]

Table XX shows the relationship between American Standard Speed, S, and American Standard Exposure Indices, Z. The numerical scales are those specified in the American Standard mentioned above. It has been customary in this country to express the speed of negative materials in arithmetic terms. Since the photographic public is so thoroughly accustomed to using speed numbers in the arithmetic form, it seems undesirable to depart from this practice.

For a purely logarithmic method of computation of camera exposure, it is necessary to convert the values of Z to a logarithmic form. In the last column

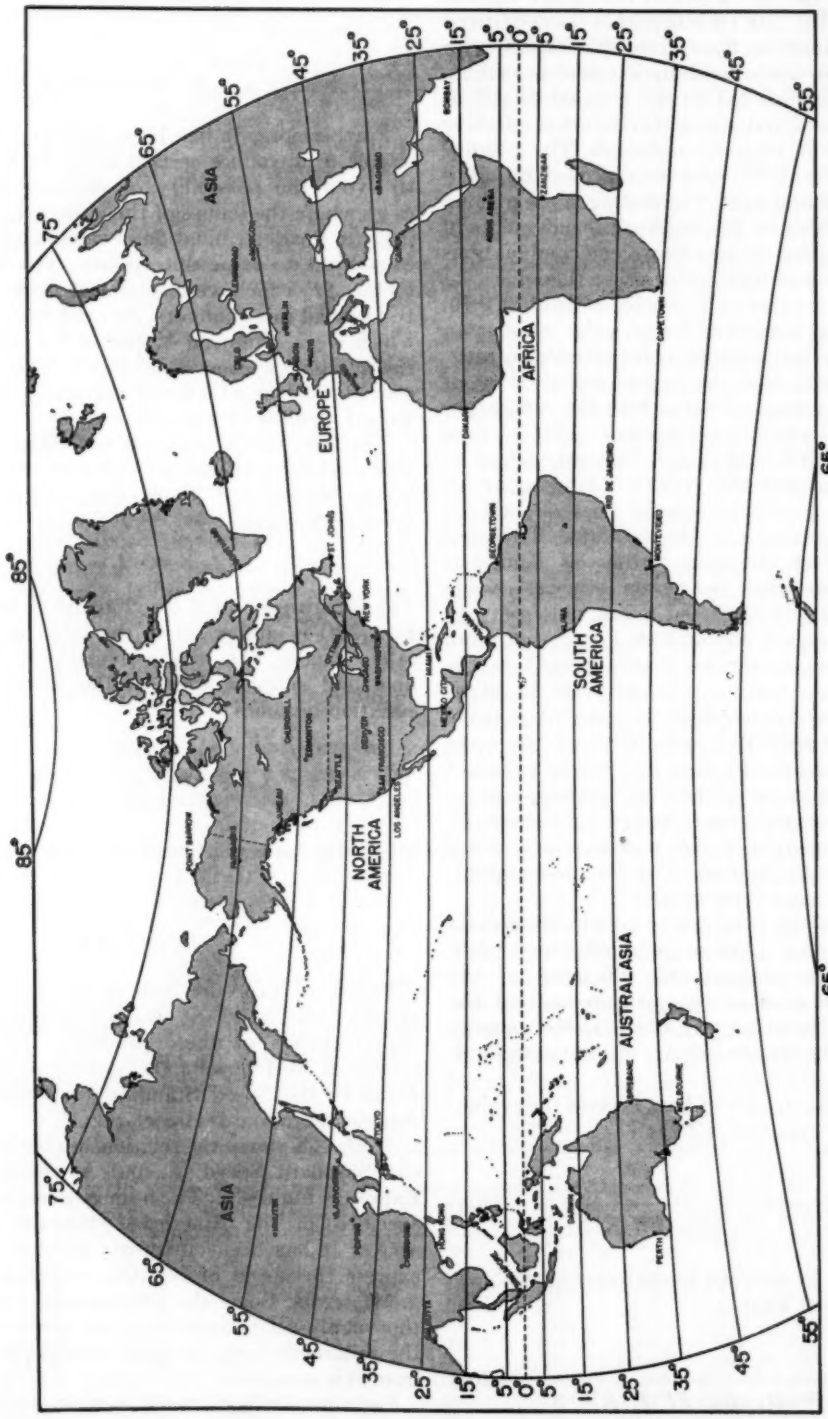


FIG. 47. Latitude zone map.

of Table XX this has been done and the values of

$$LZ = 10 \log Z \tag{45}$$

are shown.

These logarithmic equivalents should not be confused with the Logarithmic Exposure Index values given in the ASA Speed Standard as permissible for use in countries where it has been customary in the past to express the speed of photographic materials on a logarithmic scale. The values of LZ are given in this communication only to illustrate a pure logarithmic method of computation. As will be shown later, the simplest and most practicable method of computing camera exposure involves a combination of arithmetic and logarithmic numbers with a simple form of slide rule. Hence, it is quite unnecessary and undesirable for general use to adopt the LZ values shown in Table XX.

It is not within the scope of this communication to give comprehensive information concerning the exposure indices of the very numerous negative materials which are now commercially available. However, simply for the purpose of illustration, the names of a few photographic materials manufactured by the Eastman Kodak Company are given in the first column of Table XX.

The only unknown left in the second member of Eq. (20g) is B_o min. Hence, it only remains to utilize the data contained in previous sections of this paper to find the value of B_o min. for a particular condition, including: (a) the scene type (involving both aspects of scene structure, namely, *spatial distribution* and *local illumination*); (b) the *atmospheric condition;* (c) the *light value* (which depends upon the hour of the day, the month, and the geographical latitude); and (d) the existing aspect of directional lighting.

A. Arithmetic Method

The desired value of B_o min. for a particular condition may be determined by purely arithmetic operations which may be more meaningful and more easily comprehended (if not more easily accomplished) by some who are not familiar with the use of logarithmic methods. This method involves the mathematical operations of multiplication and division which may be performed easily by the use of some type of slide rule. The first step in the simplification process is the combination of the data concerning MB_o min. as dependent on scene structure, with that pertaining to atmospheric conditions. This has been done in Table XXI, from which can be read the values of MB_o min. for any scene type and for any of the atmospheric conditions which we have described. (Table XXI combines the data given previously in Tables XIV and XVII.)

The MB_o min. values contained in the body of Table XXI represent the maximal values of minimum scene luminance for the atmospheric and scene type conditions indicated, when the sun is at the zenith and apply, therefore, only when R_v and R_d are a minimum. To obtain B_o min. for other values of R_v and R_d, MB_o min. must be divided by the product of R_v and R_d. This gives

$$B_o \text{ min.} = \frac{MB_o \text{ min.}}{R_v \cdot R_d}. \tag{46}$$

On substituting this value of B_o min. in Eq. (20g), the basic arithmetic equation for camera exposure now becomes

$$\frac{t}{f^2} = \frac{R_v \cdot R_d}{MB_o \text{ min.}} \cdot \frac{1}{8.4Z}, \tag{20h}$$

where R_v = luminous density ratio Table X;[bb] R_d = directional lighting ratio, Table XVIII; Z = exposure index, Table XX.

In general, it is unnecessary to determine the value of camera exposure, t/f^2, since this is only a means to an end, the end being to find a convenient combination of t and f values which will give the

TABLE XXIII. Values of AM. AM = atmospheric index, A, + scene index, M. Solar altitude = 90°.

Atmospheric condition	A	Scene structure: Distant Un.	Distant Lt.	Distant Hv.	Semidistant Un.	Semidistant Lt.	Semidistant Hv.	Near Un.	Near Lt.	Near Hv.	Close Un.	Close Lt.	Close Hv.
Clear	0	0	3	6	3	6	9	6	9	12	9	12	15
Light haze	1.5	1.5	4.5	7.5	4.5	7.5	10.5	7.5	10.5	13.5	10.5	13.5	16.5
Medium haze	3	3	6	9	6	9	12	9	12	15	12	15	18
Heavy haze / Light cloud	5	5	8	11	8	11	14	11	14	17	14	17	20
Medium cloud	7	7	10	13	10	13	16	13	16	19	16	19	22
Heavy cloud	10	10	13	16	13	16	19	16	19	22	19	22	25
Dense cloud	15	15	18	21	18	21	24	21	24	27	24	27	30

Note: For the *local illumination* condition of *dense shade*, add 3 to the corresponding value for *heavy shade*.

[bb] Since the arithmetic method turns out to be too clumsy and laborious for practical use, we have not given tables of R_v values for latitudes other than 40°.

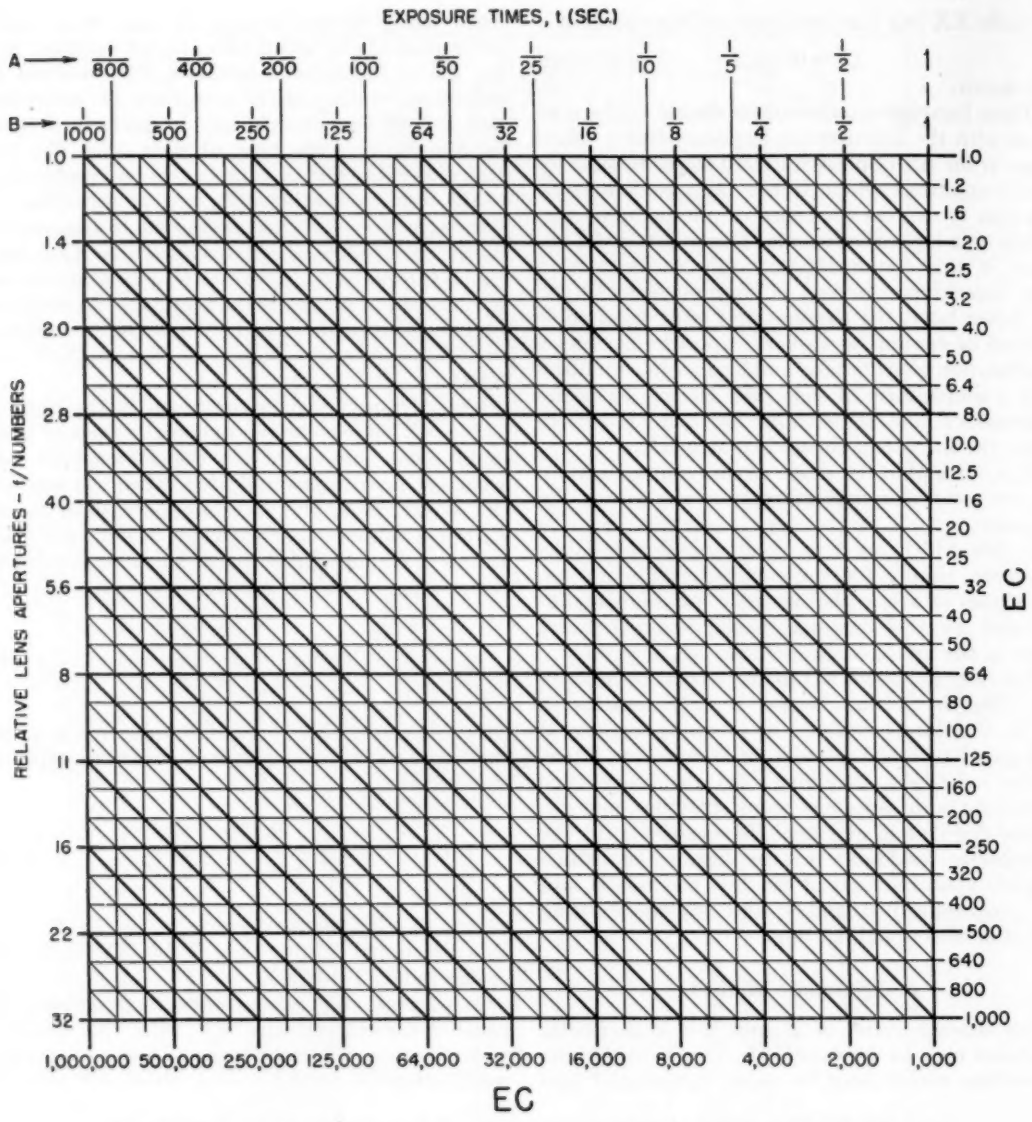


FIG. 48. Chart for the determination of correct camera exposure by a purely arithmetic method.

required camera exposure. It should be remembered that the dimension of camera exposure is *time*, and that CE may be regarded as exposure time expressed in seconds when $f=1.0$. Thus, if a table were made containing values of CE corresponding to various pairs of t and f values, the body of the table would contain numbers varying from a maximum of 1 to a minimum of 0.000001. Decimal numbers are as a rule inconvenient to use. Hence, by constructing Table XXII on the basis of the reciprocal of CE, $1/CE$, or f^2/t, whole numbers are obtained as shown. We may now write the camera exposure equation, (20h), in the reciprocal form:

$$EC=\frac{1}{CE}=\frac{f^2}{t}=\frac{MB_o \text{ min.}}{R_r \cdot R_d}\cdot 8.4Z. \qquad (47)$$

Along the top of the Table XXII are shown values of exposure time expressed (a) in milliseconds (0.001), and (b) in fractions of one second. Exposure time increases progressively from left to right by consecutive powers-of-2 steps, the fractional equivalents being shown directly below in the line labeled *b*. This series is a rational one but, unfortunately, the

values in this series do not in all cases coincide with the exposure times provided by commercially available shutters. Interpolation between the time values shown in Table XXII is therefore in some cases necessary. This cannot be done very conveniently by use of this table as constructed, but it can be accomplished more easily by use of a chart to be described later.

In the first column of Table XXII are the values of aperture ratio, f, these being in consecutive square-root-of-2 steps which are in accordance with the American Standard covering the diaphragm markings for photographic objectives.[67]

As indicated by the caption of Table XXII, the values in the body of the table are those of the reciprocals of camera exposure. They vary from 1 in the extreme upper right-hand corner of the table to 1,000,000 in the extreme lower left-hand corner.

The arithmetic solution now consists in multiplying the value of MB_0 min. (from Table XXI) by $8.4Z$, Z being the value of exposure index taken from Table XX. This product must then be divided by the product of R_v (taken from the middle section of Table X) and R_d (values which are found in Table XVIII). Now in Table XXII the number that is most nearly equal to that obtained by solving Eq. (47) is located. This indicates the correct f-value to be used with the corresponding exposure time.

We shall illustrate this process by an example. Assume that the scene to be photographed may be described as a near, unshaded scene with clear atmospheric conditions. The time of day is noon in June and the location is Washington, D. C. The scene is front-lighted and the photographic material to be used is Kodak Panatomic-X Sheet Film. Reference to Fig. 47 shows that Washington, D. C., lies within the 35°–45° latitude zone and, as indicated in Table VIII, the luminous density values, R_v, for latitude 40° therefore apply. Substituting in Eq. (47) the numerical values taken from Tables XXI, X, XVIII, and XX, we have

$$EC=\frac{f^2}{t}=\frac{16\cdot 8.4\cdot 32}{1\cdot 1}=4300. \qquad (48)$$

Assuming a value of $f=8$, and solving for t, we have

$$t=1/67,$$

indicating that the exposure time should be 1/67 second at $f/8$.

For the purpose of simplification, we shall now round off the value of 8.4 appearing in Eqs. (47) and (48) to 8.0 which will change the value 4300 in Eq. (48) to 4096. The number in Table XXII nearest to 4096 is 4000, and hence the table calls for 1/64 second at $f/8$ rather than 1/67. The change due to rounding off is less than 5 percent and this is considered negligible. We shall, therefore, in further use of Eq. (47) substitute 8.0 for 8.4, a constant derived by the use of the three values inserted in Eq. (20e). This leads to the final arithmetic equation:

$$EC=\frac{f^2}{t}=\frac{MB_a \text{ min.}}{R_v\cdot R_d}\cdot 8Z. \qquad (47a)$$

We have mentioned previously the difficulty of interpolating between the rather large steps shown in Table XXII. If a table were made in which values of EC increase by consecutive cube-root-of-2 steps, it would be necessary to use nine times as many numbers. Such a table would be clumsy and somewhat confusing to use. It is possible that a chart such as that shown in Fig. 48 may be preferred to a table expanded to the consecutive $2^{\frac{1}{3}}$ basis. Along the top of the chart are shown exposure times, t, (seconds) expressed as fractions. The heavy vertical lines are drawn at the shutter markings in row B corresponding to a consecutive powers-of-2 scale. The light vertical lines indicate the position of the intermediate $2^{\frac{1}{3}}$ steps in exposure time. In row A are located the shutter markings which are specified in the proposed American Standard for between-the-lens shutters. At the left are values of relative lens apertures corresponding to the diaphragm markings specified in the proposed American Standard covering the time marking for between-the-lens shutters. The numbers at the right and along the bottom of the chart represent values of EC. A number which has been found by the solution of Eq. (47a) is located on the EC scale, the diagonal line beginning at that point is followed to the intersection with the desired shutter setting (exposure time, t), and the horizontal line through this point is then followed to the f-scale at the left of the chart.

TABLE XXIV. Values of $X=(10\cdot\log EC)$.

Aperture ratio	(a)	Exposure times, t, in (a) milliseconds (0.001 sec.) and (b) fractions of 1 sec.										
f	(b)	1/1000	1/500	1/250	1/125	1/64	1/32	1/16	1/8	1/4	1/2	1
1		30	27	24	21	18	15	12	9	6	3	0
1.4		33	30	27	24	21	18	15	12	9	6	3
2		36	33	30	27	24	21	18	15	12	9	6
2.8		39	36	33	30	27	24	21	18	15	12	9
4		42	39	36	33	30	27	24	21	18	15	12
5.6		45	42	39	36	33	30	27	24	21	18	15
8		48	45	42	39	36	33	30	27	24	21	18
11		51	48	45	42	39	36	33	30	27	24	21
16		54	51	48	45	42	39	36	33	30	27	24
22		57	54	51	48	45	42	39	36	33	30	27
32		60	57	54	51	48	45	42	39	36	33	30

[67] "American standard for focal lengths of lenses: marking", *American Standards Association Z38.4.4* (1942), approved August 4, 1942.

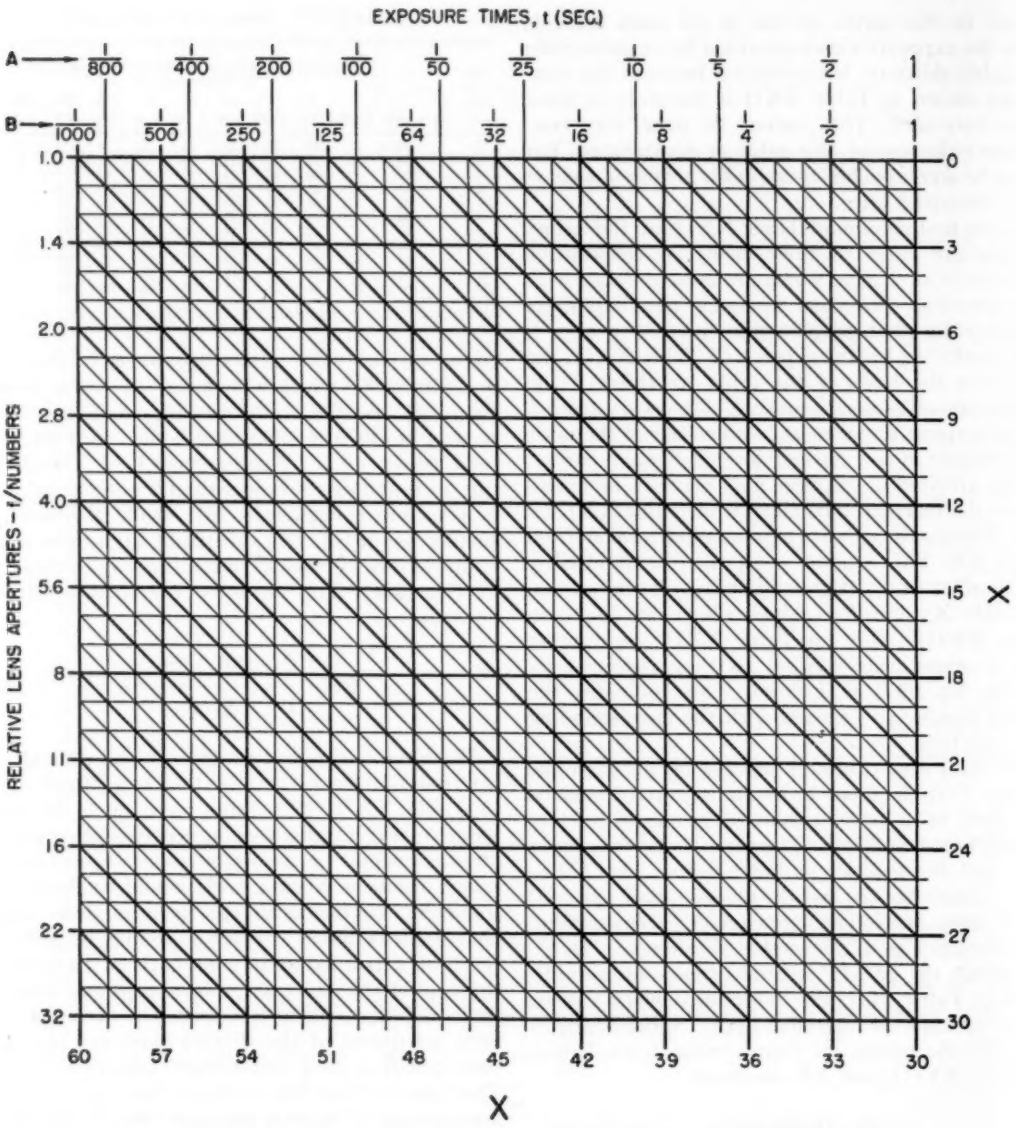


FIG. 49. Chart for the determination of correct camera exposure by a purely logarithmic method.

B. Logarithmic Method

The application of the logarithmic method involves the mathematical operations of addition and subtraction. As stated previously, these operations, provided they involve numbers consisting of not more than 2 digits, are relatively simple to perform by the average person. For the application of this method, it is necessary first of all to construct a table analogous to Table XXI based upon the values of atmospheric index, A, and scene index, M. This operation is extremely simple since it merely involves the addition of the values of A and M and the construction of the table as shown in Table XXIII. The numbers in the body of this table we shall refer to as AM. The next step is the construction of the table analogous to Table XXII but containing logarithmic values of the reciprocal of camera exposure. Thus, as indicated by the caption of Table XXIV, the numbers in the body of the table are equal to 10 logEC. We shall now convert Eq. (47a) to a logarithmic form, but before doing so we shall insert the value of MB_0 min.,

which is given by the equation,

$$MB_s \text{ min.} = \frac{64}{R_a \cdot R_s}. \tag{49}$$

This leads to

$$EC = \frac{64}{R_a \cdot R_s \cdot R_v \cdot R_d} = 8.0 \cdot Z. \tag{50}$$

This may now be written in a logarithmic form as follows:

$$\begin{aligned} 10 \log EC = {} & 10 \log 64 + 10 \log 8 + 10 \log Z \\ & - (10 \log R_a + 10 \log R_s \\ & + 10 \log R_v + 10 \log R_d). \end{aligned} \tag{51}$$

We have already defined various indices as follows:

10 log exposure index, Z, $= LZ$ (Table XX)
$10 \log (R_a + R_s) = AM$ (Table XXIII)
$10 \log R_v$ = light index, L_v (Table XII)
$10 \log R_d$ = directional index, N (Table XVIII)
$10 \log EC = X$.

Equation (51) may now be simplified to

$$X = 27 + LZ - (AM + L_v + N). \tag{52}$$

By solving Eq. (52) by the insertion of numerical values from the tables as indicated, a number is obtained which may now be taken directly to Table XXIV and the desired values of t and f obtained. A numerical example may clarify this operation and we shall assume the same scene and conditions as were specified in the numerical example given in the section on the arithmetic method:

$$X = 27 + 15 - (6 + 0 + 0) = 36.$$

It will now be seen that for a diaphragm setting of $f/8$ an exposure time of 1/64 second is called for, which is the same answer obtained by the arithmetic method.

Here, again, the possible necessity of interpolating between the powers-of-2 steps used in making up Table XXIV may indicate that a chart, such as is shown in Fig. 49, may be preferable. In this case also, the exposure-time values shown in row A represent the proposed American Standard markings and those in row B the consecutive powers-of-2 steps. The chart does make it possible to interpolate more easily and hence to obtain somewhat more precise values for f and t required to give the computed camera exposure.

C. Combined Arithmetic, Logarithmic, and Slide Rule Method

We have been unable to reduce the arithmetic method of computing camera exposure to a point

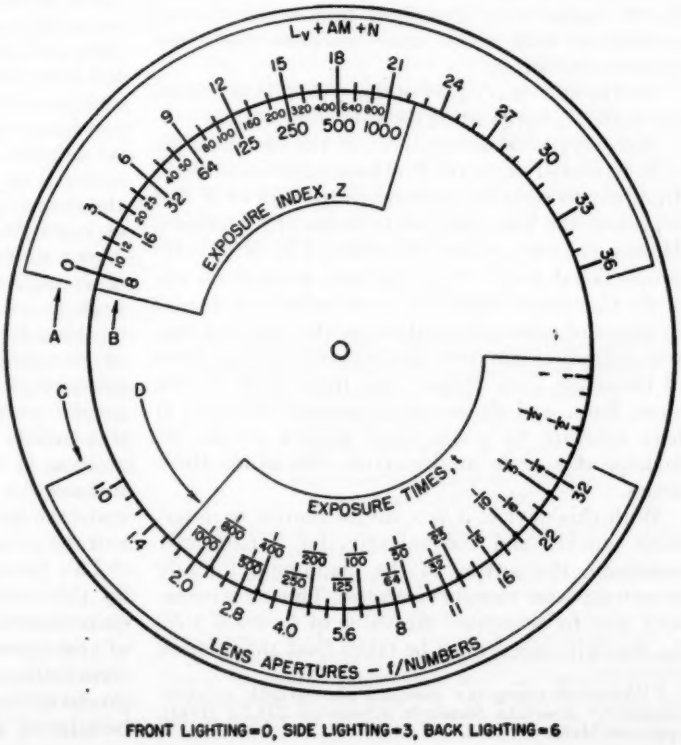


FIG. 50. Illustrates the structure of a sample circular slide rule for the determination of correct camera exposure using a combination of arithmetic and logarithmic numbers. (See American Emergency Standard, Photographic Exposure Computer, *ASA Z38.2.2*—1942.)

where it appears to have much merit for practical application. The logarithmic computational method is somewhat simpler, but still involves reference to at least four different tables, and, finally, the use either of a table such as No. XXIV or of a chart such as that shown in Fig. 49.

We have not found a method for carrying out the necessary computations which is simpler than that employed in the American Emergency Standard Photographic Exposure Computer.[68] As mentioned very early in this communication, this was formulated during the war as an emergency standard. It is probable that, with some revision, it will be issued as the American Standard. The method in question involves a combination of the arithmetic and logarithmic methods, together with the use of a simple slide rule. The form of this slide rule is shown in Fig. 50. The central portion is printed on a circular disk which may be revolved about the pivot, O. This rotating disk carries the scale, B, on which are printed the exposure indices, Z, in arithmetic cube-root-of-2 increments. The rotatable element also carries the scale, D, on which are marked the exposure times, t, in fractions of one second. The lower row of exposure times marked on the scale increase from left to right by consecutive powers of 2. In addition, the exposure times specified by the proposed American Standard for engraving on between-the-lens shutters are indicated in the upper row. The short lines which do not correspond with either scales complete the $2^{\frac{1}{3}}$ exposure-time series.

On the stationary part of the slide rule is printed the scale, A, representing the values of L_v+AM+N.

Also on the stationary part of the circular slide rule is printed the scale, C. These numbers increase from left to right by consecutive powers of $2^{\frac{1}{2}}$ and represent the lens aperture in terms of f/numbers. However, since camera exposure, CE, is directly proportional to f^2, the f/numbers marked on the scale, C, actually represent consecutive powers-of-2 in terms of exposure incident on the negative material. Each of the intervals between the long lines of the scale, C, is divided into three parts by the short lines, and these now represent increases in lens aperture by consecutive powers of the $2^{\frac{1}{3}}$, making this scale analogous to the other three scales.

With this device, it is a simple matter to determine quickly and conveniently, for a particular condition, the setting of the diaphragm for any chosen exposure time, or vice versa. Thus, it is necessary first to determine the value of L_v+AM+N; L_v, the light index, must be taken from that section of Table XII corresponding to the geographical latitude at which the work is being done; AM represents the combined atmospheric-scene index and can be found in Table XXIII. The value of N, the directional index, can be either remembered or the simple table at the bottom of the circular slide rule can be used as a memory aid. Now it is only necessary to set the exposure index of the photographic material, Scale B, at the value on Scale A equal to L_v+AM+N. The scales, C and D, are now in such a position that the f/number for any exposure time or the exposure time for any desired f/number can be read directly. This computation can be carried out in much less time that it takes to describe the technique and has proved from years of experience to be quite satisfactory and simple enough for practical use.

Thus far we have been unable to find a method which eliminates the necessity of referring to the various sections of Table XII for obtaining the value of the light index, L_v. Nomographs have been developed which give the correct answer, but none has been found which is simple enough to be better, in our opinion, than the reference to the individual latitude sections of Table XII.

XII. CONCLUSION

In bringing this communication to a close, it is desirable to make a few remarks which perhaps may serve to crystallize in the mind of the reader our aims and the consequences of the method which has been followed. We have endeavored to discuss and to evaluate all of those factors which must be considered in computing the camera exposure to be administered to a sample of photographic material in the camera for obtaining a negative of the desired quality from the standpoint of correct exposure. We have attempted to present in a useful form a summary of a large number of experimental observations which have been made by many workers on the amount of sunlight and skylight reaching the earth's surface. For the clear atmospheric condition, these data are of relatively high precision, certainly of sufficient precision for photographic purposes. To obtain a number specifying this factor, that is, the volume density of light arriving at the earth's surface, it is necessary only to know the solar altitude, which of course depends upon the hour of the day, the month of the year, and the geographical latitude.

This basic value must of necessity be modified by the existing atmospheric conditions. We are quite confident that a careful study of our discussion of atmospheric condition will enable an intelligent conscientious worker to estimate, with sufficient precision, the modification which must be applied because of the presence of various amounts and

[68] "American emergency standard photographic exposure computer," *American Standards Association Z38.2.2* (1942), approved March 26, 1942.

distributions of cloud in the earth's atmosphere. Probably the only other item with respect to which some uncertainty may be encountered is the judgment of scene type. Here, again, we feel that the verbal descriptions and pictorial illustrations which have been given of different scene types should enable one to obtain a sufficiently reliable estimate of scene type and, consequently, of scene index.

In the final equation for the determination of correct camera exposure, certain safety factors have been incorporated which, in our opinion, are of sufficient magnitude to take care of possible uncertainties in the estimates of the values of the various factors involved in the problem. For convenience, this final equation is repeated:

$$CE=\frac{t}{f^2}=\frac{1}{8.4B_o\,\text{min.}\cdot Z}. \qquad (20g)$$

The application of this equation, which contains a safety factor of 4.0, should, *on the average*, result in a value of camera exposure which is approximately four times that of the *least* camera exposure required to yield a negative from which an *excellent* print can be made. We have made a careful study of the magnitudes of the errors which may arise in making the estimates of the values of atmospheric and scene indices and of the possible departures from our assumed average values for the other factors involved in the development of Eq. (20g). The maximum total error in the value of camera exposure resulting from the application of Eq. (20g) will, of course, be encountered when all of the errors are either positive or negative. We are confident that in only a very small percentage of cases will a value of camera exposure be obtained which is *less* than the *least*, or, on the other hand, greater than *eight times the least* which is required to yield a negative from which an excellent print can be made. If *all* of the various factors which contribute to the final value of camera exposure were known with *certainty*, it would be unnecessary to employ a safety factor. It is quite possible, for instance, to measure the value of B_o min. with a photometer (free from flare light), which obviates the necessity of the worker making an estimate of the atmospheric condition and scene type. It is possible to calibrate photographic shutters and diaphragms with high precision. It is generally not possible in the field to measure the effective value of flare factor. This value is dependent not only upon the camera-lens assembly (which, of course, can be calibrated from the standpoint of flare light) but also upon the distribution of luminance within the picture area and within the environment of that picture area. It will be recalled that a partial safety factor ($SF'=2.0$) was used to take care of the unavoidable variability of the effective flare factor. Hence, even if all factors, with the exception of the flare factor, were known with precision, it would be unwise to use a safety factor of less than 2. It follows, therefore, that, assuming precise knowledge of those conditions amenable to *premeasurement* which control the amount of light reaching the negative material in the camera, it would be safe to give about one-half of the camera exposure called for by Eq. (20g).

We emphasize again that our treatment of the subject applied primarily only to the black-and-white negative-positive process and to the use of photographic negative materials for this purpose. It is well known that other photographic processes, such as black-and-white and color reversal processes, exhibit much less latitude with respect to camera exposure and hence do not permit the use of a safety factor as great as that we have employed. The sensitivity of the photographic materials has been expressed in terms of *exposure indices* determined according to the American Standard, the scope of which is limited to photographic materials "intended for the making of monochromatic continuous-tone negatives in pictorial photography." As yet no nationally recognized standard method of specifying exposure indices for reversal processes has been established. Likewise, no such standard method for measuring exposure indices for color processes of the negative-positive type has been adopted. This does not mean that our treatment of the exposure problem is entirely without validity for the purpose of determining correct camera exposure for reversal processes or for color negative-positive processes. However, the evaluation of exposure indices for such materials and processes must be made in a manner which will fit in with the requirements of Eq. (20g). A discussion of the determination of such exposure indices is beyond the scope of this communication.

Errata: Sunlight and Skylight as Determinants of Photographic Exposure. I. Luminous Density as Determined by Solar Altitude and Atmospheric Conditions

LOYD A. JONES AND H. R. CONDIT
Communication No. 1155 from the Kodak Research Laboratories, Eastman Kodak Company, Rochester, New York
[J. Opt. Soc. Am. 38, 123 (1948)]

IN the caption to Fig. 17, page 151, "($FF=2.0$)" should read "($FF=4.0$)."

In the caption to Fig. 22, page 168, "Qualitative representative" should read "Qualitative representation."

In Table XIII, page 170, the heading of the fourth column should read "Sky luminance" instead of "Sky illuminance."

In Table XIV, page 174, the number shown in the second row of the fourth column should read "1.5" instead of "15."

In reference No. 35, page 178, "A. Benporad" should read "A. Bemporad."

JOURNAL OF THE OPTICAL SOCIETY OF AMERICA VOLUME 39, NUMBER 2 FEBRUARY, 1949

Analysis of Elliptical Polarization

M. RICHARTZ
Catholic University, Peiping, China

AND

HSIEN-YÜ HSÜ
Yenching University, Peiping, China

(Received September 28, 1948)

In order to analyze an elliptical light wave, we must find its state of polarization which is completely determined by three parameters, i.e., the direction (azimuth), the form (ellipticity), and the sense of the ellipse described by the light vector. In this paper a survey is given of the direct and indirect methods used for the measurement of the magnitudes of the three parameters. First, a mathematical treatment, which uses the calculus of quaternions, of the state of polarization for special cases will give the background to the theories of the methods discussed.

I. INTRODUCTION

IN a recent paper, H. G. Jerrard[1] gave a detailed survey of the present-day knowledge of optical compensators. He limited his discussion of elliptical polarization to the measurement of phase differences only. In many cases, however, the determination of the phase difference of an elliptical wave is not sufficient. The general problem of elliptical polarimetry consists of finding the position, the form, and the sense of the ellipse described by the light vector, whereupon the state of polarization of the light is completely determined.

In this paper we shall give a complete survey of the most useful methods for analyzing elliptical polarization. The position of the elliptical vibration is given by its azimuth, i.e., by the angle ϑ ($0 \leqq \vartheta \leqq \pi$) between the major axis of the ellipse and a given fixed direction F (see Fig. 1). The form of the ellipse results from the ellipticity, i.e., from the ratio $r_2/r_1 = \tan\chi$ of the axes of the ellipse. The sense of describing the ellipse or the direction of rotation of the light vector depends upon the sign of the phase lead ψ of the x component of the vibration over the y component. Thus, the experimental analysis of elliptically polarized light consists of three distinct parts; the determining of the azimuth, the measuring of the ellipticity, and the finding of the direction of rotation. Methods which directly determine these three parameters employ an analyzing prism with half-shadow device for the determination of azimuth and add a compensating device for measuring the ellipticity, while the direction of rotation is found in considering the position of the compensating system. Indirect methods use the relation of the azimuth and of the ellipticity with the phase difference and the ratio of the amplitudes of the incident elliptical wave.

Before going into discussion, it may be a good help to investigate mathematically the state of polarization produced by a doubly refracting plate for various positions.

II. MATHEMATICAL BACKGROUND

In his survey of optical compensators, Jerrard developed the related mathematical theory following the classical method of treatment. Accordingly, he made constant references to the lengthy intensity formulas in order to calculate the half-shadow position and sensitivity of the various compensators.

We note, however, that since 1941 a new formulation of the mathematical theory of a general optical system has been furnished by R. Clark Jones,[2] and that, following his matric theory, we have been able to find not merely the magnitude, which is proportional to the intensity, but also directly the electric vector of the light which comes out through a sequence of retardation plates.

In what follows we endeavor to adopt the matric theory to the study of elliptical polarization and propose to supply a new, simpler method of computation in obtaining under various conditions the electric components of the light which characterize the state of polarization.

[1] H. G. Jerrard, J. Opt. Soc. Am. 38, 35 (1948).

[2] R. Clark Jones, J. Opt. Soc. Am. 31, 488 (1941).

1. Quaternion and Retardation Plate

The effect of a crystalline plate in the state of polarization of the light passing through it has been considered and shown by Jones[3] as to be always representable mathematically by the product of three two-by-two complex matrices. Recalling the notations and parameters as used by Jerrard (see Fig. 2), we may put the electric vector of the light after passing through a retardation plate as follows:

$$\begin{pmatrix} E_x \\ E_y \end{pmatrix} = \begin{pmatrix} \cos\gamma & -\sin\gamma \\ \sin\gamma & \cos\gamma \end{pmatrix} \begin{pmatrix} e^{i\rho} & 0 \\ 0 & e^{-i\rho} \end{pmatrix} \times \begin{pmatrix} \cos\gamma & \sin\gamma \\ -\sin\gamma & \cos\gamma \end{pmatrix} \begin{pmatrix} E_{x0} \\ E_{y0} \end{pmatrix}, \quad (1)$$

where $2\rho=\delta$. Or, we may write it symbolically as

$$\mathcal{E}=\mathbf{S}(\gamma)\mathbf{G}(\rho)\mathbf{S}(-\gamma)\mathcal{E}_0, \quad (2)$$

where the product

$$\mathbf{M}\equiv\mathbf{S}(\gamma)\mathbf{G}(\rho)\mathbf{S}(-\gamma) \quad (3)$$

acts as the transformation of the electric vector and represents the effect of a retardation plate to the incident light.

In a previous publication,[4] we have introduced, for the sake of convenience, certain new notations

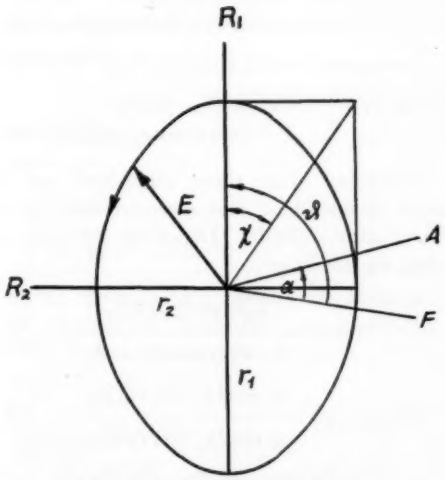


FIG. 1. Ellipse described by the light vector E; OF, fixed direction; OA, vibration direction of analyzer; α, its azimuth; OR_1, OR_2, direction of axes of the ellipse; r_1, r_2, magnitude of major and minor semiaxes; ϑ, azimuth of the major axis; $r_2/r_1=\tan\chi$, ellipticity.

[3] See reference 2, p. 490.
[4] H. Y. Hsü, M. Richartz, and Y. K. Liang, J. Opt. Soc. Am. 37, 99 (1947).

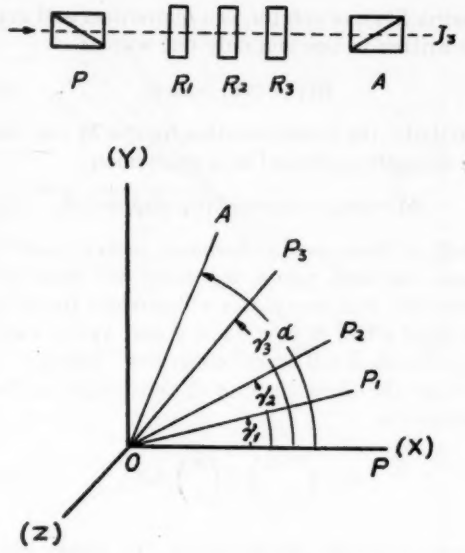


FIG. 2. Optical system and azimuths of the directions of vibration for a system of three doubly refracting plates. A, analyzer; P, polarizer; P_1, P_2, P_3, "fast" directions of vibration of the three plates; α, γ_1, γ_2, and γ_3, azimuths of the analyzer and the plates 1, 2, and 3, respectively, measured from OP.

in order to separate the real and imaginary parts in **M**. It seems to us more convenient in the course of actual multiplication to go one step further to introduce the notion of a quaternion. We denote

$$\mathbf{I}=\begin{pmatrix} 1 & 0 \\ 0 & 1 \end{pmatrix}; \quad \mathbf{i}=\begin{pmatrix} i & 0 \\ 0 & -i \end{pmatrix}; \quad \mathbf{j}=\begin{pmatrix} 0 & 1 \\ -1 & 0 \end{pmatrix}; \quad \mathbf{k}=\begin{pmatrix} 0 & i \\ i & 0 \end{pmatrix}, \quad (4)$$

where **I** is an idemfactor. It is easy to show that

$$\mathbf{i}^2=\mathbf{j}^2=\mathbf{k}^2=-1,$$

$$\mathbf{jk}=\mathbf{i}=-\mathbf{kj}; \quad \mathbf{ki}=\mathbf{j}=-\mathbf{ik}; \quad \mathbf{ij}=\mathbf{k}=-\mathbf{ji}. \quad (5)$$

Then a quaternion is a linear function of **i**, **j**, **k**,

$$\mathbf{q}=t+x\mathbf{i}+y\mathbf{j}+z\mathbf{k}, \quad (6)$$

where t, x, y, z are real numbers. Analogous to the property of conjugate complex numbers, the conjugate of **q** is defined by

$$\bar{\mathbf{q}}=t-x\mathbf{i}-y\mathbf{j}-z\mathbf{k}, \quad (7)$$

and its norm is

$$\mathbf{q}\bar{\mathbf{q}}=t^2+x^2+y^2+z^2. \quad (8)$$

For our present consideration, the rotation

matrix $\mathbf{S}(\gamma)$, or rotator, is a quaternion and may be written in one and only one way as

$$\mathbf{S}(\gamma)=\cos\gamma-\sin\gamma\mathbf{j}. \tag{9}$$

Similarly, the transformation matrix $\mathbf{M}$ can also be uniquely expressed as a quaternion:

$$\mathbf{M}=\cos\rho+\sin\rho\cos2\gamma\mathbf{i}+\sin\rho\sin2\gamma\mathbf{k}. \tag{10}$$

Both of them are, furthermore, unitary quaternions, for both norms are unity. We note also from (10) that two plates will produce the same physical effect if $\rho_2=-\rho_1=\rho$ and $\gamma_2=\gamma_1+\pi/2=\gamma$. We shall call them "congruent" plates.

Since the electric vector $\mathcal{E}_0$ emerging from the polarizer is

$$\mathcal{E}_0=\begin{pmatrix}E_{x0}\\E_{y0}\end{pmatrix}=\begin{pmatrix}1\\0\end{pmatrix}Ae^{i\omega t}, \tag{11}$$

where A is the amplitude of the plane light wave, we have

$$\mathcal{E}_1=\mathbf{M}_1\mathcal{E}_0=\begin{pmatrix}\cos\rho_1+i\sin\rho_1\cos2\gamma_1\\ i\sin\rho_1\sin2\gamma_1\end{pmatrix}Ae^{i\omega t}. \tag{12}$$

Referring to the axis of the analyzer, we have

$$\mathcal{E}_{1(a)}=\mathbf{S}(-\alpha)\mathcal{E}_1, \tag{13}$$

and the electric vector emerging from the analyzer is

$$\mathcal{E}_a'=\begin{pmatrix}E_{x,1(a)}\\0\end{pmatrix},$$

where

$$E_{x,1(a)}=(1\ 0)\mathbf{S}(-\alpha)\mathcal{E}_1=\{\cos\rho_1\cos\alpha+i\sin\rho_1\cos(\alpha-2\gamma_1)\}Ae^{i\omega t}. \tag{14}$$

The intensity emerging from the analyzer is

$$\begin{aligned}J_1&=KE_{x,1(a)}\overline{E_{x,1(a)}}\\&=\cos^2\alpha-\sin^2\rho_1(\cos^2\alpha-\cos^2(\alpha-2\gamma_1))\\&=\cos^2\alpha+\sin^2\delta_1/2\sin2\gamma_1\sin2(\alpha-\gamma_1),\end{aligned} \tag{15}$$

whereas the intensity of the light emerging from the polarizer is taken as unity.

We observe that the above method of derivation from the electric vector to the intensity of the light, the procedure of which is essentially self-explanatory in nature, will be advantageously employed in our subsequent considerations.

2. Equivalence Theorem

With the retardation plate interpreted as a quaternion, we shall readily show the first equivalence theorem of Hurwitz and Jones.[5] Our proof as given below is rather of a more elementary character. Since

$$\mathbf{S}(\gamma_1)\mathbf{S}(\gamma_2)=\mathbf{S}(\gamma_1+\gamma_2), \tag{16}$$

we shall state without loss of generality the equivalence theorem as follows: Any optical system which contains two retardation plates is optically equivalent to a system containing only one plate succeeded by a rotator. Symbolically, we want to prove

$$\mathbf{M}^{(2)}=\mathbf{M}_2\mathbf{M}_1=\mathbf{M}\mathbf{S}(-\gamma'), \tag{17}$$

that is, to find the values of ρ, γ in $\mathbf{M}$ and γ' in $\mathbf{S}(-\gamma')$ expressed uniquely as functions of ρ_2, γ_2, ρ_1, γ_1 in $\mathbf{M}_2$ and $\mathbf{M}_1$. Expressing $\mathbf{M}_2$, $\mathbf{M}_1$ and $\mathbf{M}$, $\mathbf{S}$ in quaternion as (10) and (9), by actual multiplication of the first pair $\mathbf{M}_2$ and $\mathbf{M}_1$ as well as the second $\mathbf{M}$ and $\mathbf{S}$ in accord with the rules of quaternion products as given in (5), and then equating the components, we have

$$\cos\rho\cos\gamma'=\cos\rho_2\cos\rho_1-\sin\rho_2\sin\rho_1\times\cos(2\gamma_2-2\gamma_1), \tag{18.1}$$

$$\sin\rho\cos(2\gamma+\gamma')=\cos\rho_2\sin\rho_1\cos2\gamma_1+\sin\rho_2\cos\rho_1\cos2\gamma_2, \tag{18.2}$$

$$\cos\rho\sin\gamma'=\sin\rho_2\sin\rho_1\sin(2\gamma_2-2\gamma_1), \tag{18.3}$$

$$\sin\rho\sin(2\gamma+\gamma')=\cos\rho_2\sin\rho_1\sin2\gamma_1+\sin\rho_2\cos\rho_1\sin2\gamma_2. \tag{18.4}$$

We remember that these equations are not entirely independent, but are connected by the unitary norm relation. Denoting $x=\tan\rho$ and solving, we easily get

$$\tan\gamma'=\frac{x_2x_1\sin(2\gamma_2-2\gamma_1)}{1-x_2x_1\cos(2\gamma_2-2\gamma_1)}, \tag{19.1}$$

$$\tan(2\gamma+\gamma')=\frac{x_1\sin2\gamma_1+x_2\sin2\gamma_2}{x_1\cos2\gamma_1+x_2\cos2\gamma_2}, \tag{19.2}$$

$$x^2=\tan^2\rho=\frac{x_1^2+2x_1x_2\cos(2\gamma_2-2\gamma_1)+x_2^2}{1-2x_1x_2\cos(2\gamma_2-2\gamma_1)+x_1^2x_2^2}. \tag{19.3}$$

The uniqueness of the solution can easily be seen. For, let ρ_0, γ_0, and γ_0' be a second set of

[5] H. Hurwitz, Jr. and R. Clark Jones, J. Opt Soc. Am. 31, 493 (1941).

solution, we shall have by (18.1)–(18.4)

$$\cos\rho_0\cos\gamma_0' = \cos\rho\cos\gamma',$$
$$\sin\rho_0\cos(2\gamma_0+\gamma_0') = \sin\rho\cos(2\gamma+\gamma'),$$
$$\cos\rho_0\sin\gamma_0' = \cos\rho\sin\gamma',$$
$$\sin\rho_0\sin(2\gamma_0+\gamma_0') = \sin\rho\sin(2\gamma+\gamma').$$

Eliminating, we get

$$\tan\gamma_0' = \tan\gamma', \quad \tan(2\gamma_0+\gamma_0') = \tan(2\gamma+\gamma').$$

Since the azimuth of the "fast" axis of the plate is measured between 0 and π, thus

$$\gamma_0' = \gamma'; \quad \gamma_0 = \gamma \text{ or } \gamma+\pi/2.$$

Consequently, as the phase difference $-\pi<\delta<\pi$, and hence $-\pi/2<\rho<\pi/2$, we have

$$\rho_0 = \rho \text{ when } \gamma_0 = \gamma,$$

and

$$\rho_0 = -\rho \text{ when } \gamma_0 = \gamma+\pi/2.$$

They are "congruent" plates. Therefore the equivalence theorem is completely established.

3. Analysis of Elliptic Light Emerging from M

It is well known that rotators can only change the orientation but not the shape of the ellipse representing the state of polarization of the light. We shall give a brief account of the effect of the plane light wave from the polarizer passing through **M**.

For the discussion of elliptic light, we consider the expression

$$x = a\cos\omega t, \quad y = b\cos(\omega t-\psi). \tag{20}$$

It represents in general an ellipse, whose explicit equation is

$$\frac{x^2}{a^2\sin^2\psi} - \frac{2xy\cos\psi}{ab\sin^2\psi} + \frac{y^2}{b^2\sin^2\psi} = 1. \tag{21}$$

Hence we have, for the azimuth ϑ of the major axis of the ellipse, measured counter-clockwise from the x axis, and for the ratio, $r_2/r_1=\tan\chi$, of axes of the ellipse:[6]

$$\tan 2\vartheta = \frac{2ab\cos\psi}{a^2-b^2}, \quad \sin 2\chi = \mp\frac{2ab\sin\psi}{a^2+b^2}. \tag{22}$$

The state of polarization of the light is completely determined if we know the azimuth ϑ, the ratio $\tan\chi$, and the sense of describing the ellipse.

In dealing with our case of a single retardation

[6] Max Born, *Optik* (Verlagsbuchhandlung Julius Springer, Berlin, 1933), p. 23.

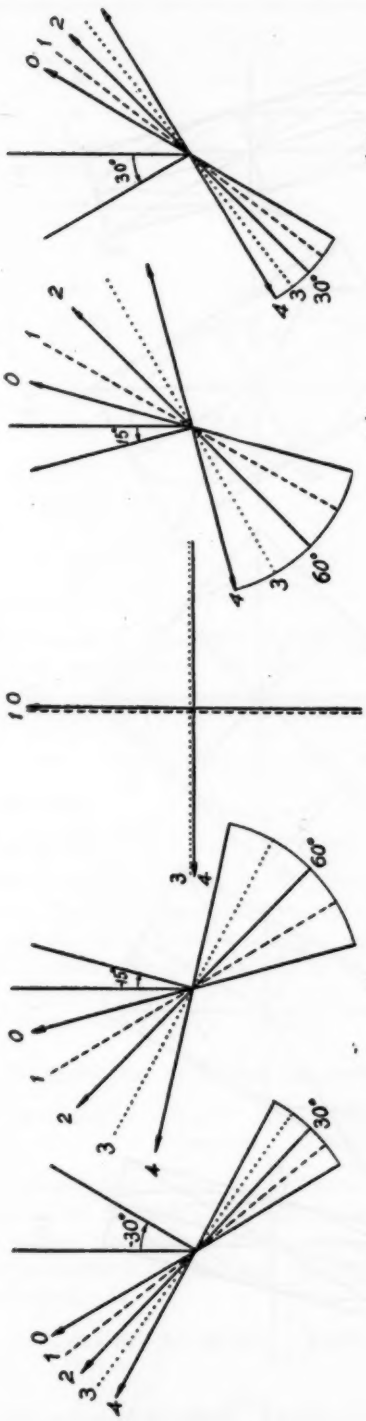


FIG. 3. Directions of the major axis, ϑ', for different $\gamma'(a-e)$; numbers 0, 1, 2, 3, 4, indicate the values of $\rho=0$, $<\pi/4$, $\pi/4$, $>\pi/4$ and $\pi/2$, respectively.

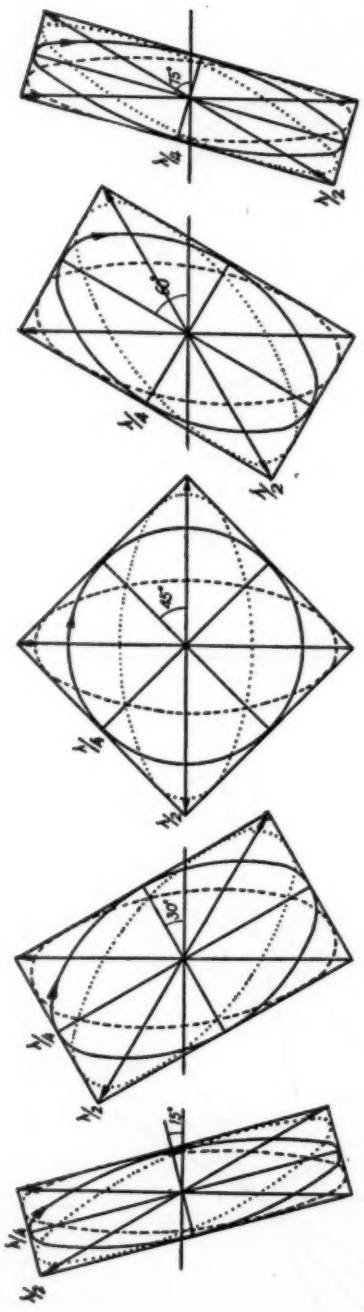


FIG. 4. Ellipses $\mathcal{E}_1$ which are $\mathcal{E}^*(\rho, \gamma')$ together with the rotator $\mathbf{S}(\gamma')$ as fore multiplier. The values of γ and ϑ correspond to the values of γ' and ϑ' in Fig. 3.

plate **M**, we shall first put for the sake of symmetry [see Eq. (12)]

$$\mathcal{E}_1=\mathbf{M}_1\mathcal{E}_0=\mathbf{S}(\gamma')\mathbf{S}(\pi/4)\mathbf{G}(\rho)\mathbf{S}(-\pi/4)\times\mathbf{S}(-\gamma')\begin{pmatrix}1\\0\end{pmatrix}Ae^{i\omega t}$$

$$=\mathbf{S}(\gamma')\begin{pmatrix}\cos\rho\cos\gamma'-i\sin\rho\sin\gamma'\\-\cos\rho\sin\gamma'+i\sin\rho\cos\gamma'\end{pmatrix}Ae^{i\omega t}\equiv\mathbf{S}(\gamma')\mathcal{E}^*(\rho,\gamma')Ae^{i\omega t},\quad(23)$$

where $\gamma'=\gamma-\pi/4$. Apart from the rotator $\mathbf{S}(\gamma')$, we find at once that

$$a^2=\cos^2\rho\cos^2\gamma'+\sin^2\rho\sin^2\gamma',$$
$$b=\cos^2\rho\sin^2\gamma'+\sin^2\rho\cos^2\gamma'.$$

And, as in (20), denoting ψ to be the difference of the arguments of y and x components of the electric vector in consideration, we obtain

$$\tan2\vartheta'=-\sec2\rho\tan2\gamma',$$
$$\sin2\chi=\mp\sin2\rho\cos2\gamma',\quad(24)$$

where $\vartheta'=\vartheta-\gamma'$. We remark that in the second equation when ρ is small, χ is also small; thus $|2\chi|=|2\rho\cos2\gamma'|=|\delta\sin2\gamma|$ is known as the effective phase difference of δ. We further derive

$$r_2/r_1=\tan\chi=\left[\frac{1-(1-\sin^2 2\rho\cos^2 2\gamma')^{\frac{1}{2}}}{1+(1-\sin^2 2\rho\cos^2 2\gamma')^{\frac{1}{2}}}\right]^{\frac{1}{2}}\equiv F^*(\rho,\gamma'),\quad(25)$$

where r_1 and r_2 are taken to be major and minor semiradii of the ellipse, respectively, and

$$r_1^2+r_2^2=a^2+b^2=1.$$

We may accordingly tabulate the results of azimuth and ellipticity of the elliptically polarized light for various values of ρ and γ (or γ') in Table I.

Figures 3 and 4 illustrate Table I for $0<\gamma<\pi/2$. For values of γ, $\pi/2<\gamma<\pi$, we have exactly the same figures except that the sense of the ellipse is in every case reversed. We note that it is a trivial case when $\rho=0$, namely a plain plate; it is denoted by the number 0 in the figures. When $\rho=\pi/2$, the plate acts towards the plane polarized light emerging from the polarizer like a rotator, whose angle of rotation is twice the azimuth of the plate [see Eq. (12) and the number 4 in the figures]. Limiting cases occur when $\rho=\pi/4$ (see Figs. 3 and 4, c). For values of ρ, $\pi/2<\rho<\pi$, we use "congruent" plates. For

TABLE I. Azimuth ϑ' (and ϑ), electric vector $\mathcal{E}^*(\rho, \gamma')$, and ellipticity $F^*(\rho, \gamma')$ for given values of ρ and γ' (or γ).

	γ γ'	0 $-\pi/4$	$0<\gamma<\pi/4$ $-\pi/4<\gamma'<0$	$\pi/4$ 0	$\pi/4<\gamma\pi/2$ $0<\gamma'<\pi/4$	$\pi/2$ $\pi/4$
	ϑ'	$\pi/4$	$0<\vartheta'<\pi/4$	0	$-\pi/4<\vartheta'<0$	$-\pi/4$
	ϑ	0	$0<\vartheta<\pi/4$	0	$-\pi/4<\vartheta<0$	0
$0<\rho<\pi/4$	$\mathcal{E}^*(\rho, \gamma')$	$\begin{pmatrix}1/\sqrt{2}\\1/\sqrt{2}\end{pmatrix}$	$\mathcal{E}^*(\rho, \gamma')$	$\begin{pmatrix}\cos\rho\\i\sin\rho\end{pmatrix}$	$\mathcal{E}^*(\rho, \gamma')$	$\begin{pmatrix}1/\sqrt{2}\\-1/\sqrt{2}\end{pmatrix}$
	$F^*(\rho, \gamma')$	$0/1$	$F^*(\rho, \gamma')$	$\dfrac{\sin\rho}{\cos\rho}$	$F^*(\rho, \gamma')$	$0/1$
	ϑ'	$\pi/4$	$\pi/4$	$\sim$	$-\pi/4$	$-\pi/4$
	ϑ	0	$0<\vartheta<\pi/4$	$\sim$	$-\pi/4<\vartheta<0$	0
$\rho=\pi/4$	$\mathcal{E}^*(\rho, \gamma')$	$\begin{pmatrix}1/\sqrt{2}\\1/\sqrt{2}\end{pmatrix}$	$\mathcal{E}^*(\pi/4, \gamma')$	$\begin{pmatrix}1/\sqrt{2}\\i/\sqrt{2}\end{pmatrix}$	$\mathcal{E}^*(\pi/4, \gamma')$	$\begin{pmatrix}1/\sqrt{2}\\-1/\sqrt{2}\end{pmatrix}$
	$F^*(\rho, \gamma')$	$0/1$	$F^*(\pi/4, \gamma')$	$\dfrac{1/\sqrt{2}}{1/\sqrt{2}}$	$F^*(\pi/4, \gamma')$	$0/1$
	ϑ'	$\pi/4$	$\pi/4<\vartheta'<\pi/2$	$\pi/2$	$\pi/2<\vartheta'<3\pi/4$	$3\pi/4$
	ϑ	0	$0<\vartheta<\pi/2$	$\pi/2$	$\pi/2<\vartheta<\pi$	π
$\pi/4<\rho<\pi/2$	$\mathcal{E}^*(\rho, \gamma')$	$\begin{pmatrix}1/\sqrt{2}\\1/\sqrt{2}\end{pmatrix}$	$\mathcal{E}^*(\rho, \gamma')$	$\begin{pmatrix}\cos\rho\\i\sin\rho\end{pmatrix}$	$\mathcal{E}^*(\rho, \gamma')$	$\begin{pmatrix}1/\sqrt{2}\\-1/\sqrt{2}\end{pmatrix}$
	$F^*(\rho, \gamma')$	$0/1$	$F^*(\rho, \gamma')$	$\dfrac{\cos\rho}{\sin\rho}$	$F^*(\rho, \gamma')$	$0/1$

the intermediate values of ρ, namely $0<\rho<\pi/4$ and $\pi/4<\rho<\pi/2$, we choose a certain fixed value of ρ in each range and draw in Fig. 3 the values of ϑ' when γ' takes the values of -30, -15, 0, $+15$, $+30$ degrees. Figure 4 shows the ellipses for the corresponding γ to be equal to 15, 30, 45, 60, and 75 degrees. We observe that for all values of ρ, $0\leqq\rho\leqq\pi/2$, the azimuths ϑ' and ϑ lie in a definite sector of 30 degrees when $\gamma=15°$ or 75°, and of 60 degrees when $\gamma=30°$ or 60°. Furthermore, for $0<\rho<\pi/4$ and when γ increases, the azimuth ϑ will gradually reach a maximum and then becomes zero when $\gamma=\pi/4$; for greater values of γ the azimuth ϑ will reach a minimum and then becomes zero again for $\gamma=\pi/2$ (see Table I and Fig. 4). The extreme values of the angle ϑ may readily be found and they are such when γ' satisfies the condition

$$\tan 2\gamma' = \pm(\cos 2\rho)^{\frac{1}{2}}.$$

On the other hand, for $\pi/4<\rho<\pi/2$ the major axis of the ellipse rotates from zero to π when γ takes the values from zero to $\pi/2$.

The form of the ellipse for a fixed value of ρ varies also with the azimuth γ of the retardation plate. When γ increases from zero to $\pi/4$, the ellipse is getting fat, i.e., r_1 becomes diminishing and r_2 increasing. When γ varies from $\pi/4$ to $\pi/2$, the ellipse becomes slender, i.e., r_1 increasing and r_2 diminishing till again linearly polarized.

The sense of describing the ellipse may be obtained directly from the electric vector; for instance, in the case (see Table I)

$$\mathcal{E}^*(\rho, \gamma') = \begin{pmatrix}\cos\rho\\i\sin\rho\end{pmatrix},$$

we may put

$$E_x = A\cos\rho\cos\omega t, \quad E_y = A\sin\rho\cos(\omega t+\pi/2).$$

When $t=0$, $E_x=A\cos\rho$, $E_y=0$; when $\omega t=\pi/2$, $E_x=0$, $E_y=-A\sin\rho$. Therefore, as t increases we get the sense of description as indicated in Fig. 4. Other cases can similarly be worked out, or they will follow by a continuous transformation for various values of ρ and γ'.

4. Application to Special Compensators

a. The Szivessy "Simple Half-Shadow" Compensator

Corresponding to the diagrammatic arrangement of the Szivessy's compensator in Jerrard's paper, we may immediately write down the electric vector, before incident upon the analyzer, symbolically as

$$\mathcal{E}_2 = \mathbf{M}^{(2)}\mathcal{E}_0 = \mathbf{M}_2\mathbf{M}_1\begin{pmatrix}1\\0\end{pmatrix}Ae^{i\omega t},$$

$$\mathcal{E}_1 = \mathbf{M}_1\mathcal{E}_0 = \mathbf{M}_1\begin{pmatrix}1\\0\end{pmatrix}Ae^{i\omega t},$$

where

$$\mathbf{M}_2=\mathbf{S}(h)\mathbf{G}\left(\frac{\eta}{2}\right)\mathbf{S}(-h)=\cos\frac{\eta}{2}+\sin\frac{\eta}{2}\cos 2h\mathbf{i}+\sin\frac{\eta}{2}\sin 2h\mathbf{k}$$

and

$$\mathbf{M}_1=\mathbf{S}\left(\pm\frac{\pi}{4}\right)\mathbf{G}\left(\frac{\delta}{2}\right)\mathbf{S}\left(\mp\frac{\pi}{4}\right)=\cos\frac{\delta}{2}\pm\sin\frac{\delta}{2}\mathbf{k}.$$

Therefore,

$$\mathcal{E}_2=\begin{pmatrix}\cos\frac{\eta}{2}\cos\frac{\delta}{2}\mp\sin\frac{\eta}{2}\sin\frac{\delta}{2}\sin 2h+i\sin\frac{\eta}{2}\cos\frac{\delta}{2}\cos 2h\\ \pm\sin\frac{\eta}{2}\sin\frac{\delta}{2}\cos 2h+i\left\{\sin\frac{\eta}{2}\cos\frac{\delta}{2}\sin 2h\pm\cos\frac{\eta}{2}\sin\frac{\delta}{2}\right\}\end{pmatrix}Ae^{i\omega t},$$

and

$$\mathcal{E}_1=\begin{pmatrix}\cos(\delta/2)\\ \pm i\sin(\delta/2)\end{pmatrix}Ae^{i\omega t}.$$

Closely following the method as described in (II, 1) and substituting $\alpha=\pi/2$, we find for the electric vector emerging from the analyzer

$$E_2=E_{x,2(a)}=\left\{\pm\sin\frac{\eta}{2}\sin\frac{\delta}{2}\cos 2h+i\left(\sin\frac{\eta}{2}\cos\frac{\delta}{2}\sin 2h\pm\cos\frac{\eta}{2}\sin\frac{\delta}{2}\right)\right\}Ae^{i\omega t},$$

and

$$E_1\equiv E_{x,1(a)}=\pm i\sin(\delta/2)Ae^{i\omega t};$$

thus,

$$J_2=\sin^2(\delta/2)+\sin^2(\eta/2)\cos\delta\sin^2 2h\pm\tfrac{1}{2}\sin\delta\sin\eta\sin 2h,$$

and

$$J_1=\sin^2(\delta/2).$$

The half-shadow azimuth h_0 of the compensator plate will satisfy $J_2-J_1=0$, namely:

$$\sin 2h_0=0,\quad\text{or}\quad\tan\delta=\mp\tan\frac{\eta}{2}\sin 2h_0. \tag{26}$$

The half-shadow sensitivity s which is defined as

$$s=\frac{1}{J_0}\left|\frac{\partial}{\partial h}(J'-J'')\right|_{h=h_0}$$

may thereby be computed. Or, we may express s entirely in terms of E_2 and E_1, and we have in our present case:

$$s=\frac{1}{E_1\bar{E}_2}\left|\frac{\partial}{\partial h}(E_2\bar{E}_2-E_1\bar{E}_1)\right|_{h=h_0}$$

$$=\frac{2}{J_1}\left|R(E_2)\frac{\partial}{\partial h}R(E_2)+I(E_2)\frac{\partial}{\partial h}I(E_2)\right|$$

$$=2\left|\frac{\sin\eta}{\tan\delta/2}\right|\left|\left(1-\frac{\tan^2\delta}{\tan^2\eta/2}\right)^{\frac{1}{2}}\right|, \tag{27}$$

where $R(E)$ and $I(E)$ are the real and imaginary parts of E.

b. *The Brace Compensator*

In order to avoid confusion about the notation for the arrangement (A) of the Brace compensator, we introduce β', instead of β, to denote the azimuth of the polarizer referring to the analyzer and still keep β as the azimuth of the analyzer with respect to the polarizer. Thus, we note $\beta'=-\beta$. The electric vector referring to the analyzer may then be written symbolically as

$$\mathcal{E}_{3(a)}=\mathbf{S}(-\beta)\mathbf{M}^{(3)}\mathcal{E}_0=\mathbf{S}(\beta')\mathbf{M}_3\mathbf{M}^{(2)}\mathcal{E}_0$$

$$=\mathbf{S}(\beta')\mathbf{S}(k)\mathbf{G}(\kappa/2)\mathbf{S}(-k)\mathbf{M}^{(2)}\mathcal{E}_0$$

$$=\mathbf{S}(k')\mathbf{G}(\kappa/2)\mathbf{S}(-k')\mathbf{S}(\beta')\mathbf{M}^{(2)}\mathcal{E}_0$$

$$=\mathbf{M}'^{(3)}\mathbf{S}(\beta')\mathcal{E}_0.$$

Now, since $d'=h'=\pm\pi/4$, we have

$$\mathcal{E}_{3(a)}=\mathbf{M}'^{(3)}\left(\cos\frac{\eta+\delta}{2}\pm\sin\frac{\eta+\delta}{2}k\right)\begin{pmatrix}\cos\beta'\\ \sin\beta'\end{pmatrix}Ae^{i\omega t}.$$

The electric vector emerging from the analyzer is

$$\mathcal{E}_a'''=\begin{pmatrix}E_{x,3(a)}\\ 0\end{pmatrix},$$

where

$$E_{x,\,3(a)}=\Big\{\cos\beta'\Big(\cos\frac{\kappa}{2}\cos\frac{\delta+\eta}{2}\mp\sin\frac{\kappa}{2}\times\sin\frac{\delta+\eta}{2}\sin 2k'+i\sin\frac{\kappa}{2}\cos\frac{\delta+\eta}{2}\cos 2k'\Big)+\sin\beta'\Big(\mp\sin\frac{\kappa}{2}\sin\frac{\delta+\eta}{2}\cos 2k'\pm i\cos\frac{\kappa}{2}\sin\frac{\delta+\eta}{2}+i\sin\frac{\kappa}{2}\cos\frac{\delta+\eta}{2}\sin 2k'\Big)\Big\}Ae^{i\omega t}.$$

Hence,

$$J_3=\cos^2\beta'-\cos 2\beta'\Big(\sin^2\Big(\frac{\delta+\eta}{2}\Big)+\sin^2\frac{\kappa}{2}\cos(\delta+\eta)\sin^2 2k'\pm\cos\frac{\kappa}{2}\sin\frac{\kappa}{2}\times\sin(\delta+\eta)\sin 2k'\Big)+\sin 2\beta'\sin^2\frac{\kappa}{2}\sin 2k'\cos 2k'. \quad (28)$$

The intensity, J_2, for the other half of the field of view is evidently the same expression as J_3 where $\eta=0$. The half-shadow azimuth $\bar{k}'$ will satisfy $J_3-J_2=0$, namely:

$$\pm\Big(1-2\sin 2\bar{k}'\sin^2\frac{\kappa}{2}\Big)\sin\Big(\delta+\frac{\eta}{2}\Big)+\sin 2\bar{k}'\sin\kappa\cos\Big(\delta+\frac{\eta}{2}\Big)=0. \quad (29)$$

We remark that if only small values of δ are considered, then we have approximately

$$\delta=\mp\sin\kappa(\sin 2\bar{k}'-\sin 2\bar{k}_0'), \quad (30)$$

where $\bar{k}_0'$ is the new half-shadow compensator azimuth of Eq. (29) with the plate D removed, namely $\delta=0$.

For the arrangement (B) of the Brace compensator, we have

$$\mathcal{E}_{3(a)}=\mathbf{S}(-\alpha)\mathbf{M}^{(3)}\mathcal{E}_0.$$

Since $d=h=\pm\pi/4$, we may write

$$\mathcal{E}_{3(a)}=\mathbf{S}(-\alpha)\mathbf{S}\Big(\pm\frac{\pi}{4}\Big)\mathbf{G}\Big(\frac{\eta}{2}\Big)\mathbf{S}\Big(\mp\frac{\pi}{4}\Big)\times\mathbf{S}(k)\mathbf{G}\Big(\frac{\kappa}{2}\Big)\mathbf{S}(-k)\times\mathbf{S}\Big(\pm\frac{\pi}{4}\Big)\mathbf{G}\Big(\frac{\delta}{2}\Big)\mathbf{S}\Big(\mp\frac{\pi}{4}\Big)\mathcal{E}_0. \quad (31)$$

The electric vector emerging from the analyzer is

$$\mathcal{E}_a'''=\begin{pmatrix}E_{x,\,3(a)}\\0\end{pmatrix},$$

where

$$E_{x,\,3(a)}=\Big[\cos\alpha\Big(\cos\frac{\kappa}{2}\cos\frac{\delta+\eta}{2}\mp\sin\frac{\kappa}{2}\sin\frac{\delta+\eta}{2}\sin 2k+i\sin\frac{\kappa}{2}\cos\frac{\delta-\eta}{2}\cos 2k\Big)+\sin\alpha\Big(\pm\sin\frac{\kappa}{2}\sin\frac{\delta-\eta}{2}\cos 2k\pm i\cos\frac{\kappa}{2}\sin\frac{\delta+\eta}{2}+i\sin\frac{\kappa}{2}\cos\frac{\delta+\eta}{2}\sin 2k\Big)\Big]Ae^{i\omega t}.$$

Hence,

$$J_3=\cos^2\alpha-\cos 2\alpha\Big(\sin^2\frac{\delta+\eta}{2}+\sin^2\frac{\kappa}{2}\cos(\delta+\eta)-\sin^2\frac{\kappa}{2}\cos^2 2k\cos\delta\cos\eta\pm\sin\frac{\kappa}{2}\cos\frac{\kappa}{2}\sin(\delta+\eta)\sin 2k\Big)+\sin 2\alpha\Big(\sin\frac{\kappa}{2}\cos 2k\Big\{\sin\frac{\kappa}{2}\sin 2k\cos\delta\pm\cos\frac{\kappa}{2}\sin\delta\Big\}\Big). \quad (32)$$

Similarly, the intensity, J_2, is the same as J_3 with $\eta=0$. Consequently, the half-shadow azi-

muth $\bar{k}$ is given by

$$\pm\sin\delta\left(\cos\kappa\cos\frac{\eta}{3}\mp\sin 2\bar{k}\sin\kappa\sin\frac{\eta}{2}\right)$$
$$+\cos\delta\left\{\sin\kappa\sin 2\bar{k}\cos\frac{\eta}{2}\right.$$
$$\left.\pm\sin\frac{\eta}{2}\left[1-2\sin^2 2\bar{k}\sin^2\frac{\kappa}{2}\right]\right\}=0.$$

Each of the half-shadow sensitivity s_A and s_B may be immediately obtained from the definition and the corresponding expression for J_2.

III. DIRECT METHODS

A. Determination of the Azimuth

An approximate value of the azimuth ϑ can be obtained by using an analyzer only. Let the elliptically polarized wave under examination be incident on the analyzer provided with a divided circle, and let α_0 be the azimuth of the analyzer with respect to the fixed direction F when the illumination of the field of view is as weak as possible. Then α_0 gives the azimuth of the minor axis of the elliptical vibration, and the azimuth ϑ of the major axis is given by $\vartheta=\alpha_0+\pi/2$.

To find a more accurate value of ϑ, half-shadow methods are in use. With regard to the wave incident upon the analyzer, four types of half-shadow devices may be distinguished.

(1) The elliptical wave is falling unchanged on a "double-field analyzer." In the types (2) to (4) a crystal plate is introduced into the beam of the elliptically polarized light and something of the ellipse is changed before falling on the analyzer.

(2) The direction of the axes of the ellipse is changed by means of a rotating biplate.

(3) The sense of the ellipse in one half of the field is reversed by using a half-wave plate.

(4) The direction as well as the form of the ellipse is varied by a doubly-refracting biplate.

1. Double-Field Analyzer[7]

Lippich, in his polarimeter,[8] has devised a double-field polarizer which in the elliptical polarimetry is often used as double-field analyzer. A single, large analyzing prism is placed immediately behind a second smaller analyzer covering only half the field. The principal section of the smaller prism is inclined at a small angle ϵ, called half-shadow angle, to that of the larger analyzer. Lippich's arrangement has the great advantage that the half-shadow angle can be varied. But a main objection lies in the different transmission of the light for the two fields, leading to systematical errors.[9]

We have, theoretically, that the incident elliptical vibration of ellipticity $\tan\chi=\tan\rho$ at an azimuth ϑ to a fixed direction is readily given by

$$\mathcal{E}_1=\mathbf{S}(\vartheta)\begin{pmatrix}\cos\rho\\ \pm i\sin\rho\end{pmatrix}Ae^{i\omega t}.$$

According to Eq. (13), we have

$$\mathcal{E}_{1(\alpha)}=\mathbf{S}(-\alpha)\mathcal{E}_1 \quad\text{and}\quad \mathcal{E}_\alpha'=\begin{pmatrix}E_{x,\,1(\alpha)}\\ 0\end{pmatrix},$$

where

$$E_{x,\,1(\alpha)}=\{\cos\rho\cos(\alpha-\vartheta)\pm\sin\rho\sin(\alpha-\vartheta)\}Ae^{i\omega t};$$

and

$$\mathcal{E}_{2(\alpha)}=\mathbf{S}(\epsilon)\mathcal{E}_{(\alpha)}^* \quad\text{and}\quad \mathcal{E}_\alpha''=\begin{pmatrix}E_{x,\,2(\alpha)}\\ 0\end{pmatrix},$$

where $\mathcal{E}_{(\alpha)}^*$ is the same as $\mathcal{E}_\alpha'$ except for α being replaced by $\alpha+\epsilon$, and

$$E_{x,\,2(\alpha)}=\cos\epsilon\{\cos\rho\cos(\alpha-\vartheta+\epsilon)$$
$$\pm\sin\rho\sin(\alpha-\vartheta+\epsilon)\}Ae^{i\omega t}.$$

Thus,

$$J_1=\cos^2\rho\cos^2(\alpha-\vartheta)+\sin^2\rho\sin^2(\alpha-\vartheta),$$
$$J_2=\cos^2\epsilon\{\cos^2\rho\cos^2(\alpha-\vartheta+\epsilon)$$
$$+\sin^2\rho\sin^2(\alpha-\vartheta+\epsilon)\}.$$

For half-shadow position, $J_1-J_2=0$; hence, as ϵ is small but not zero, we have

$$\sin\epsilon\{\cos^2\rho\cos^2(\alpha-\vartheta+\epsilon)$$
$$+\sin^2\rho\sin^2(\alpha-\vartheta+\epsilon)$$
$$=-\sin(2\alpha-2\vartheta+\epsilon)\cos 2\rho. \quad (33)$$

It is apparent that for any given incident elliptical wave there exist no exact, simple relations between α and ϑ. However, since ϵ is small, $\sin(2\alpha-2\vartheta+\epsilon)$ must also be small. For weak illumination, we have approximately $\vartheta=\alpha+\epsilon+\pi/2$. We further note that even if the incident light is linearly polarized, the half-shadow position occurs only with the simple relation $\vartheta=\alpha_0+\epsilon$, while the illumination of the field of

[7] O. Schönrock, *Handbuch der Physik* 19, 749 (1928).
[8] F. Lippich, Wien. Ber. 91, 1059 (1885).
[9] G. Szivessy, *Handbuch der Physik* 19, 955 (footnote 2) (1928); T. M. Lowry, *Opt. Rotatory Power* (Longmans, Green and Company, London, 1935), p. 189.

view is not just as weak as the sensitivity requires. As a consequence, the Lippich double-field analyzer should not be employed for accurate measurements.

Schönrock[10] devised a double-field analyzer consisting of two equal Glan-Thomson prisms which are cemented together after a thin wedge has been removed from each of their interface. Thus, the light emerging from the two analyzers is plane polarized in directions inclined at a small angle 2ϵ to each other and being symmetrical to the bisector BB of the shadow angle (Fig. 5). While the shadow angle is fixed, the diminution of intensity caused by reflection and absorption is the same in the two prisms.

The procedure to determine the azimuth ϑ of an elliptically polarized wave is as follows. The analyzer is brought perpendicular to the normal of the wave, and the telescope is focused on the dividing line of the double-field analyzer. It needs a careful adjustment to obtain a satisfactory dividing line. The two halves of the field of view will appear equally but weakly illuminated when the bisector of the shadow angle is parallel to the minor axis of the ellipse. Let the direction of the bisector be taken as the azimuth of the double-field analyzer denoted by α, and let α_0 be its azimuth in the half-shadow position. Then $\vartheta=\alpha_0+\pi/2$. This half-shadow condition follows easily from the theory.

For Schönrock's arrangement we have

$$J_1=\cos^2\rho\cos^2(\alpha-\vartheta+\epsilon)+\sin^2\rho\sin^2(\alpha-\vartheta+\epsilon),$$
$$J_1'=\cos^2\rho\cos^2(\alpha-\vartheta-\epsilon)+\sin^2\rho\sin^2(\alpha-\vartheta-\epsilon).$$

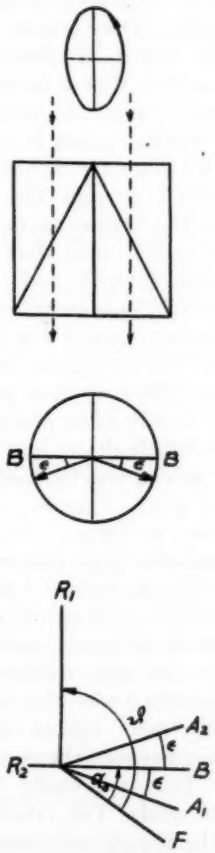


FIG. 5. The Schönrock analyzer in the half-shadow position.

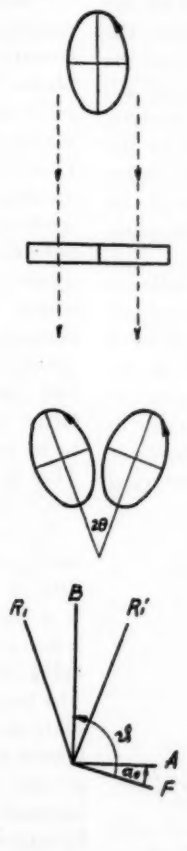


FIG. 6. The effect of the Nakamura plate upon the elliptical vibration.

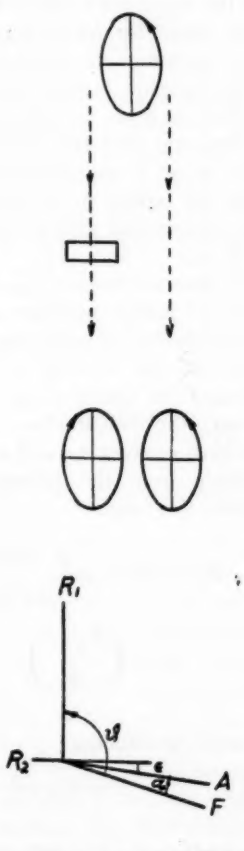


FIG. 7. The effect of the half-wave plate when its axes are parallel to that of the incident, elliptical wave.

[10] See reference 7, p. 750.

The half-shadow condition is

$$\cos 2\rho \sin(2\alpha - 2\vartheta) \sin 2\epsilon = 0. \qquad (34)$$

Thus, for weak illumination $\vartheta = \alpha_0 + \pi/2$.

2. *Rotating Biplate*[11]

A bi-quartz plate, devised by Soleil, was constructed by cementing together two semicircular disks of dextro- and laevo-quartz cut perpendicular to the optic axis having the same thickness. It was originally designed for the use of white light in a special type of saccharimeter; thus a particular thickness (3.75 mm) of the bi-quartz plate was chosen. Bertrand[12] first used such a biplate for testing elliptical polarization. Nakamura[13] showed that the biplate is very sensitive if its thickness is small (0.04 mm). In that case the rotations θ and $-\theta$ caused by the biplate are small for all frequencies. Like the Schönrock analyzer, its sensitivity cannot be varied, but the thin biplate has the great advantage that it gives a practically vanishing dividing line, and that the dividing line of the biplate can be set in any direction as we please by rotating the biplate in its plane; thus it can be used in conjunction with a half-shadow plate (see III, *B*, 1a).

The procedure for determining ϑ is as follows: The biplate is placed in front of an analyzer perpendicular to the normal of the elliptical wave in question, and the telescope is focused on the dividing line of the biplate (Fig. 6). By rotating the analyzer, the half-shadow azimuth α_0 of weak illumination gives $\vartheta = \alpha_0 + \pi/2$. This result follows simply from the following theoretical consideration. We have

$$\mathcal{E}_{1(a)} = \mathbf{S}(-\alpha)\mathbf{S}(\vartheta+\theta)\begin{pmatrix} \cos\rho \\ \pm i \sin\rho \end{pmatrix} A e^{i\omega t}$$

and

$$\mathcal{E}_a' = \begin{pmatrix} E_{x,\,1(a)} \\ 0 \end{pmatrix},$$

where

$$E_{x,\,1(a)} = \{\cos(\alpha-\vartheta-\theta)\cos\rho \pm i \sin(\alpha-\vartheta-\theta)\sin\rho\} A e^{i\omega t}.$$

Thus,

$$J_1 = \cos^2\rho \cos^2(\alpha-\vartheta-\theta) + \sin^2\rho \sin^2(\alpha-\vartheta-\theta),$$

whereas J_1' is the same as J_1 except for the angle θ being replaced by $-\theta$. From (34) the half-shadow azimuth α_0 satisfies $\vartheta = \alpha_0 + \pi/2$.

A rotator constructed in combining two half-wave plates[14] may be very useful in analyzing elliptical polarization. Two half-wave plates are combined in such a way that the "fast" axis of the one plate forms an angle ϵ with the "slow" axis of the second one. Such a combination rotates an incident (linearly or) elliptically polarized vibration through an angle 2ϵ. For this, mathematically, we have

$$\begin{aligned}\mathbf{M}_2\mathbf{M}_1 &= \{\cos 2[\gamma + (\pi/2) \pm \epsilon]\mathbf{i} \\ &\quad + \sin 2[\gamma + (\pi/2) \pm \epsilon]\mathbf{k}\} \\ &\quad \times \{\cos 2\gamma \mathbf{i} + \sin 2\gamma \mathbf{k}\} \\ &= \cos(\pm 2\epsilon) - \sin(\pm 2\epsilon)\mathbf{j} = \mathbf{S}(\pm 2\epsilon), \qquad (35)\end{aligned}$$

which shows the combination is a rotator.

Furthermore, it is readily proved that this combination is the *simplest* rotator which can be constructed by means of doubly refracting plates. For it is clear that a single retardation plate cannot be a rotator. The condition for two plates to form a pure rotator is that $\rho = 0$ (see II, 2). From (19.3) we have then either that the numerator vanishes or that the denominator tends to infinity. The former case may be possible only first when $\gamma_2 = \gamma_1$, and thus $x_1 = -x_2$; or second when $\gamma_2 = \gamma_1 \pm \pi/2$, and thus $x_1 = x_2$, which is however "congruent" to the first instance. We further observe that the condition $\gamma_2 = \gamma_1$ will render the two plates as one equivalent plate; hence this case does not hold. The latter case has already been proved from (35) to be possible, when both are half-wave plates. That it is only so can also be easily seen from (19.3) in making $\rho = 0$.

For our purpose, we form a "bi-rotator" consisting of one half-wave plate covering the whole field of view, while the second half-wave plate is a biplate, consisting of two halves each of which covers half the field only, their "fast" axes being inclined to each other through an angle ϵ. The biplate is combined with the first half-wave plate so that the "slow" axis of the first plate bisects the angle ϵ between the two "fast" axes of the biplate. That can easily be arranged between crossed nicols. The effect of such a bi-rotator may be immediately seen to be the same as that of the Nakamura plate.

It may be noted that this bi-rotator is more economical than the rotating quartz plate,

[11] See reference 9, p. 932.
[12] E. Bertrand, Bull. Soc. Mineral. **1**, 26 (1878).
[13] S. Nakamura. Centralblatt f. Min., 267 (1905).

[14] Such rotator was proposed by Hans Mueller at M.I.T. in a private conversation (July 1946).

because mica can be obtained more easily than the dextro- and laevo-quartz. Since the phase difference of a mica plate of definite thickness varies appreciably throughout the visible spectrum, a series of half-wave plates, correct respectively for the red, yellow, green, and blue regions is required. But the construction of such plates presents no problem, since mica is easily split to the thickness desired.[15]

3. *Half-Shadow Plate*[16]

Chauvin,[17] in analyzing elliptical polarization, combined the analyzer with a half-wave plate covering half the field of view. The vibration directions make an angle ϵ with each other. This combination has the advantage of variable sensitivity, but the disadvantage of different transmission of light in the two fields as the Lippich double-field analyzer has. Furthermore, the half-wave plate limits Chauvin's arrangement to the use of one particular wave-length of monochromatic light.

When an elliptical wave is incident upon the system, a half-shadow position of the analyzer can be obtained by rotating the combined analyzer and half-wave plate. We see that

$$\mathcal{E}_2=\{\cos 2(\alpha+\epsilon)\mathbf{i}+\sin 2(\alpha+\epsilon)\mathbf{k}\}\mathbf{S}(\vartheta)\begin{pmatrix}\cos\rho\\ \pm i\sin\rho\end{pmatrix}Ae^{i\omega t}$$

$$=\begin{pmatrix}\pm\sin\rho\sin(\vartheta-2\alpha-2\epsilon)+i\cos\rho\cos(\vartheta-2\alpha-2\epsilon)\\ \pm\sin\rho\cos(\vartheta-2\alpha-2\epsilon)-i\cos\rho\sin(\vartheta-2\alpha-2\epsilon)\end{pmatrix}Ae^{i\omega t}, \tag{36}$$

and

$$\mathcal{E}_{2(a)}=\mathbf{S}(-\alpha)\mathcal{E}_2, \quad \mathcal{E}_a''=\begin{pmatrix}E_{x,\,2(a)}\\ 0\end{pmatrix},$$

where

$$E_{x,\,2(a)}=\{\pm\sin\rho\sin(\vartheta-\alpha-2\epsilon)+i\cos\rho\cos(\vartheta-\alpha-2\epsilon)\}Ae^{i\omega t}.$$

Thus,

$$J_2=\sin^2\rho\sin^2(\vartheta-\alpha-2\epsilon)+\cos^2\rho\cos^2(\vartheta-\alpha-2\epsilon).$$

Corresponding to the other half-field of view, we have

$$\mathcal{E}_1=\mathbf{S}(\vartheta)\begin{pmatrix}\cos\rho\\ \pm i\sin\rho\end{pmatrix}Ae^{i\omega t}, \quad \mathcal{E}_{1(a)}=\mathbf{S}(-\alpha)\mathcal{E}_1$$

and

$$\mathcal{E}_a'=\begin{pmatrix}E_{x,\,1(a)}\\ 0\end{pmatrix},$$

where

$$E_{x,\,1(a)}=\{\cos\rho\cos(\alpha-\vartheta)\pm i\sin\rho\sin(\alpha-\vartheta)\}Ae^{i\omega t}.$$

Thus,

$$J_1=\cos^2\rho\cos^2(\alpha-\vartheta)+\sin^2\rho\sin^2(\alpha-\vartheta).$$

The half-shadow position is

$$\cos 2\rho\sin(2\alpha-2\vartheta+2\epsilon)\sin 2\epsilon=0.$$

Hence, for weak illumination the azimuth ϑ is given by $\vartheta=\alpha_0+\epsilon+\pi/2$. We note that in this half-shadow position the axes of the ellipse coincide with the principal directions of vibration of the half-wave plate. Consequently only the sense of the ellipse has been reversed (see Fig. 7) by the half-wave plate, as follows easily from the theory. For substituting $\vartheta=\alpha_0+\epsilon+\pi/2$ in (36), we obtain

$$\mathcal{E}_2=\begin{pmatrix}\pm\sin\rho\sin\vartheta-i\cos\rho\cos\vartheta\\ \mp\sin\rho\cos\vartheta-i\cos\rho\sin\vartheta\end{pmatrix}Ae^{i\omega t}$$

$$=-i\mathbf{S}(\vartheta)\begin{pmatrix}\cos\rho\\ \mp i\sin\rho\end{pmatrix}Ae^{i\omega t},$$

which shows that the sense of describing the ellipse is reversed.

Chaumont[18] used as half-shadow device a combination of a half-wave plate and of a quarter-wave plate, each covering half the field of view, but in such a way that their corresponding principal directions of vibration are crossed, and meanwhile so arranged that the phase differences in the four quadrants are: π, 0, $\pi/2$, $-\pi/2$ (see Fig. 8a). One of the principal directions is in the azimuth ϵ referred to that of the analyzer.

To find the azimuth ϑ (and the ellipticity) of elliptical polarization, this half-shadow system is combined with a quarter-wave plate covering the whole field of view. The principal directions of vibration of the quarter-wave plate coincide with that of the half-shadow device so that the

[15] J. Strong, Rev. Sci. Inst. 6, 243 (1935).
[16] See reference 9, p. 955.
[17] Chauvin, Ann. de Toulouse 3(J), 30 (1889).
[18] L. Chaumont, Ann. de physique (9), 4, 175 (1915).

phase differences in the four quadrants of this combination are: $3\pi/2$, $\pi/2$, π, 0 (see Fig. 8b). The quadrants of phase differences π and 0 work as the Chauvin half-shadow arrangement. Thus when elliptically polarized light is incident upon the system thus oriented and the whole system is rotated, the direction of the ellipse is determined by the relation $\vartheta=\alpha_0+\epsilon+\pi/2$, where α_0 is the half-shadow position of the analyzer for the two matched quadrants of the half-shadow

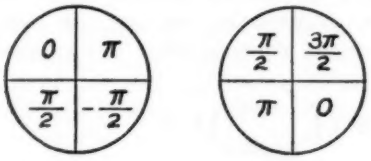


FIG. 8. Chaumont's half-shadow device. In Fig. 8a the half-wave plate covers the first and fourth quadrants with phase differences π and $-\pi/2$, the quarter-wave plate the third and fourth quadrants with phase differences $\pi/2$ and $-\pi/2$. Figure 8b shows the phase differences in the four quadrants when the half-shadow device is combined with the compensating quarter-wave plate.

device with phase differences $\pi/2$ and $-\pi/2$ (lower quadrants in Fig. 8a).

4. *Doubly Refracting Biplate*[19]

A doubly refracting biplate consists of two semicircular disks of a doubly refracting crystal plate (mica or gypsum or quartz cut parallel to the optic axis), cemented together so that their "fast" (or "slow") directions of vibration form an angle 2β to each other. To construct such a biplate, the thin crystal plate is cut in two along a direction inclined at the angle β to one principal axis of the crystal and cemented together again after turning one-half completely over.

Theory.—Let a monochromatic, elliptically polarized, plane wave be incident normally on the biplate and let the two waves emerging from the biplate traverse the analyzer. From Fig. 9 we have, mathematically, the electric vector referring to the analyzer

$$\mathcal{E}_{2(a)}=\mathbf{S}(-\alpha)\mathbf{S}(k)\mathbf{G}\left(\frac{\Delta}{2}\right)\mathbf{S}(-k)\mathbf{S}(\vartheta)\times\mathbf{S}\left(\pm\frac{\pi}{4}\right)\mathbf{G}\left(\frac{\delta}{2}\right)\mathbf{S}\left(\mp\frac{\pi}{4}\right)\mathcal{E}_0$$

$$=\mathbf{S}(-\alpha+\vartheta)\mathbf{S}(k-\vartheta)\mathbf{G}\left(\frac{\Delta}{2}\right)\mathbf{S}(-k+\vartheta)\times\mathbf{S}\left(\pm\frac{\pi}{4}\right)\mathbf{G}\left(\frac{\delta}{2}\right)\mathbf{S}\left(\mp\frac{\pi}{4}\right)\mathcal{E}_0.$$

This is, in fact, the special case of the arrangement (B) of the Brace compensator; namely, Eq. (31) when $\eta=0$. Hence, the intensity J emerging from the analyzer is

$$\begin{aligned}J=&\cos^2(\alpha-\vartheta)\\&-\cos2(\alpha-\vartheta)\left\{\sin^2\frac{\delta}{2}+\sin^2\frac{\Delta}{2}\cos\delta\right.\\&\left.\times\sin^22(k-\vartheta)\pm\tfrac{1}{2}\sin\Delta\sin\delta\sin2(k-\vartheta)\right\}\\&+\sin2(\alpha-\vartheta)\left\{\sin^2\frac{\Delta}{2}\cos2(k-\vartheta)\right.\\&\left.\times\sin2(k-\vartheta)\cos\delta\pm\tfrac{1}{2}\sin\Delta\sin\delta\cos2(k-\vartheta)\right\}.\end{aligned}$$

Let $e=\tan\delta/2=\tan\chi$, thus,

$$\sin\delta=\frac{2e}{1+e^2},\quad\cos\delta=\frac{1-e^2}{1+e^2},$$

we obtain

$$\begin{aligned}J=\frac{1}{1+e^2}\Big\{&\cos^2(\alpha-\vartheta)+e^2\sin^2(\alpha-\vartheta)\\&-(1-e^2)\sin^2\frac{\Delta}{2}\sin2(k-\vartheta)\sin2(k-\alpha)\\&\mp e\sin\Delta\sin2(k-\alpha)\Big\}.\end{aligned}$$

The intensities J' and J'' corresponding to both halves of the field of view are given by putting β_1 and $\beta_2=\beta_1+2\beta$ for the value of k, respectively. For the positive difference of J' and J'', we have

$$\begin{aligned}\delta J=|J'-J''|=\frac{1}{1+e^2}\Big|&2\sin\frac{\Delta}{2}\sin(\beta_1-\beta_2)\\&\times\left\{(1-e^2)\sin\frac{\Delta}{2}\sin2(-\alpha-\vartheta+\beta_1+\beta_2)\right.\\&\left.\times\cos(\beta_1-\beta_2)\mp2e\cos\frac{\Delta}{2}\cos(\beta_1+\beta_2-2\alpha)\right\}\Big|,\end{aligned}$$

[19] G. Szivessy and W. Herzog, Zeits. f. Instrumentenk. 58, 229, 345 (1938).

where the upper and lower sign correspond to the elliptical light emerging from the first plate to be "right" and "left" elliptical polarized, respectively (see III, C, 1). For subsequent consideration we shall, following Szivessy's work, take the upper sign.

Since $\beta_2-\beta_1=2\beta\neq 0$, there exist, according to the theory, three types of biplates in conjunction with the analyzer:

1. (Calderon) $|\beta_1-\beta_2|\neq\pi/2$; $\beta_1+\beta_2\neq\pm\pi/2$.
2. (Sommerfeldt) $|\beta_1-\beta_2|\neq\pi/2$; $\beta_1+\beta_2=\pm\pi/2$.
3. (Bravais) $|\beta_1-\beta_2|=\pi/2$; $\beta_1+\beta_2\neq\pm\pi/2$.

The Calderon plate can be used to analyze an elliptical vibration which will be discussed. The Sommerfeldt arrangement is a suitable device for determining the azimuth ϑ of an elliptical vibration. In the half-shadow position with weak illumination of the field of view, the vibration direction of the analyzer lies perpendicular to the major axis of the vibration ellipse in question. The Bravais biplate can be used only for testing elliptical polarization. It is worthy of mention that the sensitivity of the Bravais biplate is the greater, the smaller its phase difference (not quarter-wave) and the smaller the angle $(\beta_1-\alpha)$ (see Fig. 9). The Bravais biplate is employed in the method of MacCullagh-Stokes (see III, B, 3b).

Determination of ϑ with Calderon's arrangement: We take the special case that the vibration direction of the analyzer bisects the angle between the two fast directions of vibration of the biplate. Then (see Fig. 9)

$$\beta_1+\beta_2=2\alpha \quad \text{or} \quad \beta_1+\beta_2=2\alpha\mp\pi.$$

If 2β is taken small $(0<|\beta|<\pi/4)$, then four real half-shadow azimuths $\bar{\alpha}$ exist which satisfy the relation

$$\sin 2(\bar{\alpha}-\vartheta)=\pm(\tan 2\chi/\cos 2\beta)\cot(\Delta/2). \quad (37)$$

Let us choose an angle γ, defined by the equation

$$\gamma=\tfrac{1}{2}\arcsin[(\tan 2\chi/\cos 2\beta)\cot(\Delta/2)], \qquad (-\pi/4\leqq\gamma\leqq+\pi/4);$$

the half-shadow azimuths can be written:

$$\bar{\alpha}_1=\pm\gamma+\vartheta; \quad \bar{\alpha}_2=\pi/2\mp\gamma+\vartheta;$$
$$\bar{\alpha}_3=\pi\pm\gamma+\vartheta; \quad \bar{\alpha}_4=3\pi/2\mp\gamma+\vartheta.$$

Two of them, say $\bar{\alpha}_2$ and $\bar{\alpha}_4$ are of weak intensity. The corresponding readings on the graduated circle of the analyzer are t_2 and t_4.

Procedure.—As the vibration direction of the analyzer should bisect the angle 2β, we must determine the "fast" direction of vibration in the two halves of the biplate. That can be done by using a compensating glass plate subjected to a pure tensile stress. It behaves as a positive crystal with the "fast" axis in the direction of the applied stress. To adjust the biplate to the analyzer, we introduce the biplate between crossed nicols and in rotating it we determine the extinction position of one-half of the field, so that the "fast" axis of that half-plate coincides with the vibration direction of the polarizer. Then in rotating the biplate through a small angle a position will be found in which both halves of the field of view appear equally but less intensely illuminated. In this position the biplate is fastened with the analyzer. Let now the elliptically polarized wave in question be incident upon the oriented system: biplate-analyzer. Let t_0 be the analyzer reading corresponding to the analyzer azimuth $\alpha=0$. In rotating the system through 2π, we get the two half-shadow positions $\bar{\alpha}_2$ and $\bar{\alpha}_4$ with the readings

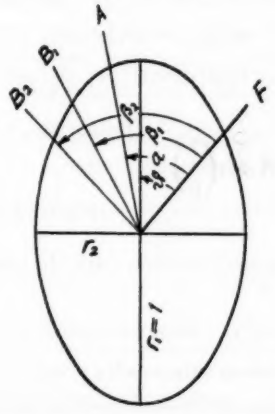


FIG. 9. Vibration directions of the combination: doubly refracting biplate and analyzer; B_1, B_2, fast axes of the half-plates 1 and 2, respectively.

t_2 and t_4. In the half-shadow position t_2 (and t_4) the analyzer is held fixed and the biplate is rotated through $+90°$ or $-90°$; the half-shadow is disturbed. It can be restored by turning biplate and analyzer together; we get the half-shadow azimuths $\bar{\alpha}_2'$ (and $\bar{\alpha}_4'$) with the corresponding readings t_2' (and t_4'). Thus ϑ is given by

$$\vartheta=(t_2+t_2')/2-(t_0+90°)=(t_4+t_4')/2-(t_0+90°).$$

Some more complicated methods for the determination of the azimuth ϑ will be discussed in connection with the measurement of the ellipticity.

B. Measurement of Ellipticity

Most methods for measuring the ellipticity are based on the following reflection. Let P denote the vibration direction of the polarizer, A the vibration direction of the analyzer being perpendicular to P. Between crossed nicols a crystal plate of unknown phase (difference δ $(0<\delta<\pi/2)$ should be introduced in azimuth $\pm\pi/4$ referred to A. The light vibration emerging from the crystal state is elliptically polarized. The major axis r_1 of the ellipse is parallel to P and the minor axis r_2 parallel to A. By means of a compensator the phase difference δ can be determined. This phase difference is obviously the difference in phase between the vibration components referred to the directions D_1 and D_2 which bisect the angles between A and P. The general relation of the ellipticity $\tan\chi$ to the phase difference δ and to the ratio $b/a=\tan\gamma$ of the amplitudes of the components of the incident wave is given by (22), namely

$$\sin 2\chi = \mp \sin 2\gamma \sin\delta. \tag{22}$$

In the above-mentioned special case, we have $\tan\gamma=1$, thus $\sin 2\gamma=1$; hence

$$\sin 2\chi = \mp\sin\delta, \quad \text{and} \quad |\chi| = \delta/2.$$

On the other hand, there are compensators, e.g., the "modified" Szivessy and the "modified" Brace compensators, which cannot determine the ratio b/a; thus the compensating method alone is not sufficient to determine the ellipticity. We shall hereafter confine ourselves to those compensators for which the ratio b/a is known and shall proceed according to the number of compensating plates used in the methods. Since a survey of the theories of optical compensators has been given by H. G. Jerrard,[1] in most cases we need give the method of use only.

There is, however, one theoretical point which seems worthy to be presently mentioned. For, in dealing with the following methods, it is desirable in several cases to make a transformation from the elliptical wave in consideration to linearly polarized light. We observe that the mathematical condition, in general, for an electric vector $\mathcal{E}$ of the light emerging from an optical system to be linearly polarized is clearly that the ratio of its x and y components should be real. For our subsequent consideration we desire to have a linearly polarized light after passage through two retardation plates still remaining to be linearly polarized. We have from (18.1)–(18.4)

$$\mathcal{E}=\mathbf{M}_2\mathbf{M}_1\begin{pmatrix}1\\0\end{pmatrix}=\begin{bmatrix}\cos\rho_2\cos\rho_1-\sin\rho_2\sin\rho_1\cos 2(\gamma_2-\gamma_1)\\ \qquad +i[\cos\rho_2\sin\rho_1\cos 2\gamma_1+\sin\rho_2\cos\rho_1\cos 2\gamma_2]\\ -\sin\rho_2\sin\rho_1\sin 2(\gamma_2-\gamma_1)+i[\cos\rho_2\sin\rho_1\sin 2\gamma_1+\sin\rho_2\cos\rho_1\sin 2\gamma_2]\end{bmatrix}.$$

Thus the required condition may be immediately obtained as

$$\frac{\cos\rho_2\sin\rho_1\cos 2\gamma_1+\sin\rho_2\cos\rho_1\cos 2\gamma_2}{\cos\rho_2\cos\rho_1-\sin\rho_2\sin\rho_1\cos 2(\gamma_2-\gamma_1)} = \frac{\cos\rho_2\sin\rho_1\sin 2\gamma_1+\sin\rho_2\cos\rho_1\sin 2\gamma_2}{-\sin\rho_2\sin\rho_1\sin 2(\gamma_2-\gamma_1)}, \tag{38}$$

an equivalent formula of which was given by Gabler and Sokob in their work on Senarmont's compensator.[20]

1. *Methods Using a Half-Shadow Plate Only*

(a) *Richartz's Method*[21]

The elliptical analyzer, devised by M. Richartz, is undoubtedly the simplest one, simple in arrangement and theory. It consists of a quarter-wave half-shadow plate H in conjunction with the Nakamura plate N (see Fig. 10). The two plates are combined in such a way that their dividing lines are crossed. Between crossed nicols, the linearly polarized light emerging from the polarizer traverses successively the half-shadow plate, the biplate and then the analyzer. In rotating the combined plates the "fast" axis of the quarter-wave plate is brought perpendicular to the vibration direction of the analyzer. In this position the two quadrants of the field of view, free from the quarter-wave plate as well as the two quadrants covered with the half-shadow plate appear equally illuminated (test for linearly polarized light). Now let the elliptical wave under examination traverse the elliptical analyzer, thus adjusted. By rotating

[20] F. Gabler and P. Sokob, Zeits. f. Instrumentenk 58, 301 (1938).
[21] M. Richartz, Zeits. f. Instrumentenkunde 60, 360 (1940).

the whole system simultaneously half-shadow appears in the two qaudrants free of the quarter-wave plate. These half-shadow positions (α_0) of the analyzer determine the azimuth of the ellipse (see III, A, 2). Since now the axes of the quarter-wave plate are parallel to that of the ellipse, the light emerging from the quarter-wave plate is linearly polarized with a vibration direction inclined to the major axis of the incident vibration. For in the present case $\rho_1=\rho$, $\gamma_1=\pm\pi/4$; $\rho_2=\pi/4$, $\gamma_2=0$ or $\pi/2$; these values satisfy the condition (38), hence the emerging light is linearly polarized. We have, accordingly,

$$\mathcal{E}=(\tfrac{1}{2})^{\frac{1}{2}}(1+i)\begin{pmatrix}\cos\rho\\ \pm\sin\rho\end{pmatrix} \text{ or } (\tfrac{1}{2})^{\frac{1}{2}}(1-i)\begin{pmatrix}\cos\rho\\ \mp\sin\rho\end{pmatrix}.$$

It inclines to the major axis of the incident vibration with the angle ρ. We can obtain half-shadow in the two other quandrants covered by the quarter-wave plate, by rotating the analyzer alone. The difference $(\alpha_1-\alpha_0)$ of the two half-shadow azimuths of the analyzer determines the ellipticity $\tan\chi=\tan(\alpha_1-\alpha_0)$.

The principal advantage of Richartz's arrangement is that in spite of the half-shadow plate the transmission of light through two quadrants when matched is exactly equal.*

(b) *Szivessy's Method*[22]

The method used in Szivessy's simple half-shade compensator for measuring very weak phase differences is employed in the determination of weak ellipticities. When the vibration direction of the analyzer has been set perpendicular to the major axis of the elliptical wave in question by means of the Nakamura plate, the biplate is replaced by the half-shadow plate. In rotating this plate we get eight half-shadow positions during one revolution, four "extinction" azimuths (used in Richartz' method), where the axes of the half-shadow plate are parallel to that of the ellipse, and four "measuring half-shadow" azimuths which play the essential role in Szivessy's method. From the latter azimuths $h_0(h_{01}, h_{02}, h_{03}, h_{04})$, corresponding to readings p_1, p_2, p_3, p_4 on the graduated circle

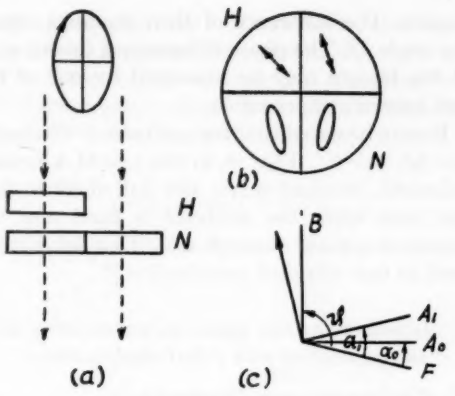


FIG. 10. Richartz' elliptical analyzer. (a) Combination of half-shadow quarter-wave plate, H, and rotating biplate, N. (b) Change of the elliptical vibration by the combination when the axes of plate H are parallel to that of the ellipse. (c) Half-shadow azimuths of the analyzer, α_0 and and α_1, for the lower and upper halves of the field of view, respectively.

of the half-shadow plate, $2h_0$ is calculated:

$$2h_0=(\theta_1+\theta_2+\theta_3)/3,$$

where $\theta_1=\pi/2-(p_2-p_1)$; $\theta_2=(p_3-p_2)-\pi/2$; $\theta_3=\pi/2-(p_4-p_3)$. The phase difference δ is then given by (26). The phase difference η of the half-shadow plate can be simply measured by using a compensating quarter-wave plate.[23] The ellipticity results from the relation $\tan\delta/2=\tan\chi$.

2. *Determination of ellipticity with Calderon's biplate*[24]

The ellipticity can be calculated from Eq. (37); thus,

$$\tan2\chi=\cos2\beta\,\tan\frac{\Delta}{2}\sin2\gamma,$$

and

$$|\chi|=\tfrac{1}{2}\arctan\left(\cos2\beta\,\tan\frac{\Delta}{2}\sin|t_2'-t_2|\right)$$

$$=\tfrac{1}{2}\arctan\left(\cos2\beta\,\tan\frac{\Delta}{2}\sin|t_4'-t_4|\right).$$

The readings t corresponding to the half-shadow azimuths have already been measured for determining the azimuth ϑ. The acute angle, 2β, can be found by determining between crossed nicols the extinction positions of the two halves of the

* It may be pointed out that Richartz' elliptical analyzer is very suitable also to measure circular dichroism. Experiments, the results of which will be published in the near future, are made in the Department of Physics at the Catholic University, Peiping.

[22] G. Szivessy and A. Dierkesmann, Ann. d. Physik **11**, 909 (1931).

[23] M. Richartz, "Measurement of phase differences of half-shadow plates," Zeits. f. Instrumentenk. **60**, 358 (1940).

[24] See reference 19, p. 345.

biplate. The difference of their readings equals the angle 2β. The phase difference Δ (about $\pi/5$) of the biplate may be measured by one of the well-known compensators.

It may be noted that for $\chi=0$ also $\gamma=0$; hence $\bar{\alpha}_2=\bar{\alpha}_2'$, $\bar{\alpha}_4=\bar{\alpha}_4'$. That is, in the case of a linearly polarized, incident wave, the half-shadow does not vary when the analyzer is fixed and the biplate is rotated through $\pi/2$. This behavior is used to test elliptical polarization.[25]

3. *Methods using two plates, a compensating plate in conjunction with a half-shadow device*

(*a*) *The Quarter-wave Compensators*

The Senarmont compensator[26] is mostly used in combination with Chauvin's half-wave plate or with Chaumont's device.

When the Chauvin analyzer is used, we first have to find the direction of the major axis of the elliptical wave in question (see III, A, 3) and thus fix the direction of the analyzer. Then the quarter-wave plate is adjusted in front of the analyzer so that its axes are parallel to that of the ellipse or, what is the same, to the axes of the half-wave plate (auxiliary adjustment between crossed nicols). When the elliptical wave traverses the quarter-wave plate, thus oriented, it is converted into linearly polarized light which is rotated by a definite angle with respect to the incident direction. By rotating the Chauvin analyzer through this angle, χ, half-shadow appears. The ellipticity is then given by $\tan\chi$.

In using Chaumont's half-shadow device we may recall that for the determination of the azimuth the quarter-wave plate in conjunction with Chaumont's device forms a Chauvin plate, and that the azimuth ϑ is given by the half-shadow azimuth (α_0) of the analyzer for the two matched quadrants of Chaumont's device with phase difference $\pi/2$ and $-\pi/2$. For the measurement of ellipticity we consider the Chaumont device combined with the analyzer as a Chauvin analyzer. Since the elliptical wave emerging from the quarter-wave plate is linearly polarized, we rotate the Chaumont analyzer until half-shadow appears in the other two quadrants, i.e., in the quadrants with phase difference 0 and π. Let α_1 be the new analyzer azimuth; then the ellipticity is given by $\tan\chi=\tan(\alpha_1-\alpha_0)$.

The Richartz quarter-wave compensators[27] can also be used to measure the ellipticity of an elliptical radiation. When the azimuth of the ellipse has been determined by means of the Nakamura plate, the half-shadow azimuth α_0 of the analyzer is known. Then the half-shadow plate and the quarter-wave plate can be brought in the azimuths required for the compensator methods; e.g., by introducing a polarizer into the elliptical wave and crossing it with the fixed analyzer. Thus the phase difference, the value of which gives twice the angle of the ellipticity, i.e., $\delta=2\chi$, can be measured.

(*b*) *Method of Mac Cullagh-Stokes*[28]

Mac Cullagh and Stokes modified the Senarmont method for the case when the phase difference of the compensating plate is not exactly but only nearly $\pi/2$. The simple arrangement consists of an analyzer and the compensating plate of phase difference δ'. Let the x direction be one of the vibration directions of the compensating plate, and ϑ be the azimuth of the elliptical, incident wave referred to the x axis. In order that the emerging light be linearly polarized, the condition (38) must be fulfilled, where $\rho_1=\chi=\delta/2$, $\gamma_1=\pm\pi/4$; $\rho_2=\delta'/2$, $\gamma_2=\vartheta$. If neither plate is quarter-wave, the condition (38) becomes

$$\tan 2\chi=\tan 2\rho_1=\mp\tan 2\rho_2 \sin 2\gamma_2 = \mp\tan\delta' \sin 2\vartheta. \quad (39)$$

Hence, there exist for either sign on the right member of (39) two positions of the compensating plate, where upon the emerging wave is linearly polarized. Let their corresponding azimuths be ϑ_1 and $\vartheta_2=\pi/2-\vartheta_1$. From the measured difference $\vartheta_2-\vartheta_1$ and the condition $\vartheta_1+\vartheta_2=\pi/2$, we get the azimuths ϑ_1 and ϑ_2 of the one axis of the ellipse in the two compensating positions.

To determine the azimuth $\bar{\vartheta}$ of the one axis of the ellipse referred to the fixed direction, analyzer and compensating plate are provided with graduated circles, each capable of rotation. Let p_0 be the reading on the graduated circle of the compensating plate and q_0 be the analyzer reading when the vibration direction (x-axis) of the compensator plate and that of the analyzer are parallel to a fixed direction; and let p_1 and q_1 be the corresponding readings in the first com-

[25] See reference 24, p. 229.
[26] See reference 9, p. 955.
[27] M. Richartz, Zeits. f. Instrumentenkunde **61**, 148 (1941); J. Opt. Soc. Am. **31**, 292 (1941).
[28] See reference 9, p. 956.

pensating position, p_2 and q_2 that in the second compensating position. Then the azimuth $\bar{\vartheta}$ results from $\vartheta_2-\vartheta_1=p_2-p_1$ and $\vartheta_1+\vartheta_2=\pi/2$ to

$$\bar{\vartheta}=(p_2-p_0)/2+(p_1-p_0)/2+\pi/4.$$

To find the compensating positions more exactly, Perucca introduced a Bravais biplate between compensator plate and analyzer in any azimuth different from 0 and $\pi/2$. Half-shadow occurs only when the light incident on the biplate is linearly polarized, that is, in the compensating positions only.

For the measurement of ellipticity, let us denote the azimuth of the emerging linearly polarized wave referred to the x axis be γ, and consider the optical system in reverse order, i.e., from linearly polarized light to elliptically polarized. Then referring to the linear vibration, the azimuth of the compensator plate is γ. Now from the analysis of elliptical light emerging from **M** (see II, 3) we have Eq. (24), namely

$$\sin 2\chi=\mp\sin\delta'\cos 2(\gamma-\pi/4)$$
$$=\mp\sin\delta'\sin 2\gamma. \quad (40)$$

Hence, there exist for a definite sign of the right member of (40) two values, γ_1 and $\gamma_2=\pi/2-\gamma_1$, corresponding to ϑ_1 and ϑ_2. From (39) and (40) we can easily obtain

$$\tan 2\vartheta=\tan 2\gamma\cos\delta', \quad \cos 2\gamma=\cos 2\chi\cos 2\vartheta. \quad (41)$$

Hence, the angle χ of the ellipticity is given by

$$\cos 2\chi=\frac{\cos 2\gamma_1}{\cos 2\vartheta_1}=\frac{\cos 2\gamma_2}{\cos 2\vartheta_2}=\frac{\sin(\gamma_2-\gamma_1)}{\sin(\vartheta_2-\vartheta_1)}$$

or

$$\cos 2\chi=\frac{\sin(q_2-q_1)}{\sin(p_2-p_1)}.$$

In using a double-field analyzer, the azimuths γ_1 and γ_2 can be determined more precisely.

More complicated in calculating the azimuth and ellipticity is the use of the half-shadow device by Tool.[29] The analyzer is a Cornu-Jellett double-field prism[30] combined with a Brace half-shadow plate, the dividing line of which is perpendicular to that of the Jellett nicol. Thus the field of view is divided into four quadrants. When a linearly or elliptically polarized wave is falling upon the Mac Cullagh-Stokes compensator, there exist two azimuths of the compensator and of the analyzer for which the four quadrants have equal intensities (compensating positions).

The theory of this arrangement was given first by Tuckerman,[31] later in a lucid form by Skinner.[32] According to Skinner, the azimuth is given by

$$\vartheta=(\alpha_1'+\alpha_2')/2+\pi/2,$$

and the ellipticity results from the equation

$$\cos 2\chi=\frac{\sin[(k_2'-k_1')-(\alpha_2'-\alpha_1')]}{\sin(k_2'-k_1')}\cdot\frac{\sin(k_2''-k_1'')}{\sin[(k_2''-k_1'')-(\alpha_2''-\alpha_1'')]},$$

where k_1', k_2' denote the azimuths of the compensator and α_1', α_2' the corresponding azimuths of the analyzer in the two compensating positions; k_1'', k_2'', α_1'', α_2'' are the azimuths of compensator and analyzer, respectively, when instead of the elliptical wave a linearly polarized wave of the same frequency is incident upon the system.

(c) *Method of the Brace Compensator*

Szivessy[33] used the Brace compensator in the arrangement A where the azimuths are taken from the vibration direction of the analyzer. After having determined the azimuth ϑ of the elliptical vibration by means of the Nakamura plate and thus fixed the direction of the analyzer, the biplate is replaced by the Brace compensator, and the half-shadow plate is given a small azimuth corresponding to desired sensitivity. Then the half-shadow azimuths k_1, k_2, k_3, k_4 are found by rotating the compensating plate through 2π. To find the half-shadow azimuths k_{01}, k_{02}, k_{03}, k_{04}, the elliptical wave is replaced by linearly polarized light without changing the position of the analyzer and the azimuth of the half-shadow plate. From the measured half-shadow azimuths k and k_0 the phase difference of the elliptical vibration is calculated from (30), namely,

$$\delta=\sin\delta'(\sin 2k-\sin 2k_0),$$

where δ' is the known phase difference of the compensating plate. Since the vibration direction of the analyzer is perpendicular to the major axis of the ellipse, we have $\chi=\mp\delta/2$.

---

[29] A. Q. Tool, Phys. Rev. **31**, 1 (1910).
[30] T. M. Lowry, *Optical Rotatory Power* (Longmans, Green and Company, London, 1935), p. 186.
[31] L. B. Tuckerman, Univ. Stud. of the University of Nebraska, **9**, 194 (1909).
[32] C. A. Skinner, J. Opt. Soc. Am. **10**, 491 (1925).
[33] G. Szivessy, Zeits. f. Instrumentenk. **47**, 148 (1925).

Bergholm[34] used Brace compensator in conjunction with a quarter-wave plate for the determination of the azimuth ϑ. Let all azimuths be referred to a fixed direction, say x axis, and let $\pi/4$ be the azimuth of the vibration direction of the analyzer. When a linearly polarized wave of the same frequency as the elliptical wave in question, and when in azimuth σ (say $\sigma=-\pi/4$), is incident upon the Brace compensator, a half-shadow azimuth $\bar{k}_0(=k_0-\pi/4)$ can be found by rotating the compensating plate. Let this half-shadow position be the zero-position H_0. Now when the elliptical wave is falling on the compensator, the half-shadow is destroyed. To restore it, the compensating plate must be rotated. We get the half-shadow azimuth $\bar{k}_0'(=k_0'-\pi/4)$. From the measured quantities $\bar{k}_0$, $\bar{k}_0'$ and the phase difference δ' of the compensating plate, we find

$$\delta=\sin\delta'(\cos2\bar{k}_0'-\cos2\bar{k}_0).$$

Since the zero-position H_0 is independent from the azimuth σ of the linearly polarized wave, the disturbance of H_0 caused by the elliptical wave can be compensated—instead of rotating the compensating plate—by changing the elliptical vibration into linearly polarized light. This reflection leads to the following determination of the azimuth ϑ. A quarter-wave plate in azimuth $\pi/4$ is placed in front of the Brace compensator into the beam of the elliptical light; then, one of its vibration directions coincides with that of the analyzer. We can easily show that the incident elliptical wave is transformed into linearly polarized light by the quarter-wave plate if and only if the axes of the ellipse coincide with the principal vibration directions of the quarter-wave plate. For we have already shown that it is linearly polarized when $\gamma_2=0$ or $\pi/2$. On the other hand, let γ_2 be kept variable, the condition (38) reduces to

$$\sin2\gamma_2\cos2\rho_1=0.$$

Hence, unless the incident light is circularly polarized, the only azimuths of the quarter-wave plate for the formation of linearly polarized light are 0 and $\pi/2$. In other words, the vibration directions of the quarter-wave plate must coincide with the axes of the elliptical vibration. This coincidence can be attained by rotating quarter-wave, Brace compensator and analyzer simultaneously through the angle $\alpha_0=\vartheta-\pi/4$. In this position again half-shadow occurs. Thus the azimuth ϑ is given by $\vartheta=\alpha_0+\pi/4$.

With measured δ and ϑ we find the ellipticity from the relation

$$\tan2\chi=\mp\sin2\vartheta\tan\delta$$

which equation may be deduced from (22).

C. Right or Left Elliptical Polarization

1. *Definitions*

In the state of elliptical light vibration the elliptic area is swept over by the light vector during one period. Let df be the element of area swept out by the light vector in the time dt; then the areal velocity is given by

$$df/dt=\tfrac{1}{2}\nu ab\sin\delta. \tag{42}$$

The areal velocity is positive or negative according as the wave normal ($+z$-axis) is circulated by the light vector in the positive or negative sense; the polarization is called "left" or "right" elliptical, respectively. Thus the polarization is right-handed when, to an observer looking in a direction opposite to that of propagation, the rotation corresponds to that of the hands of a watch. As a and b in (42) are always positive, it follows that the wave is right elliptical when $\sin\delta$ is negative, left elliptical for positive $\sin\delta$.

The phase difference δ is the difference of the phase constants between the two waves E_X and E_Y. In the case that the elliptical wave is caused by a doubly refracting plane-parallel plate, the phase difference is given by

$$\delta=2\pi(d/\lambda)(\mu_Y-\mu_X),$$

where μ_X and μ_Y are the refractive indices of rays having vibrations along OX and OY, respectively. The sign of the birefringence depends upon the kind of crystal plate. Positive uniaxial crystals have positive birefringence, $n_e>n_o$, where n_o is the refractive index of the ordinary ray, and n_e is the extreme value of the refractive index of the extraordinary ray. The ordinary ray, therefore, travels faster than the extraordinary ray; the optic axis is the "fast" direction. In negative uniaxial crystals ($n_e<n_o$) the optic axis is the "slow" direction. If we relate the azimuth of a crystal plate to that of the direction of the optic axis, the phase difference can be positive and negative. But δ remains always positive if by the azimuth of a crystal plate is meant the

[34] C. Bergholm, Physik. Zeits. 21, 137 (1920).

azimuth of the main direction of vibration of the "faster" ray. To analyze a given elliptical wave it is sufficient to consider the phase differences between 0 and 2π; when $0<\delta<\pi$, the wave is left elliptical; when $\pi<\delta<2\pi$, the wave is right elliptical.

In the applications negative phase differences are used to compensate positive δ; then the azimuth is referred to the "slow" axis of the compensating plate. In this case the wave is right elliptical when $-\pi<\delta<0$, left elliptical when $0<\delta<+\pi$.

2. *Determination of the Sense*

To find the sense of rotation of the light vector the relation (24) between phase difference and ellipticity is mostly used; namely,

$$\sin 2\chi = \mp \sin 2\gamma \sin\delta. \qquad (24)$$

When the ellipticity $\tan\chi$ is always taken positive ($0\leqq\chi\leqq\pi/4$), $\sin 2\chi$ is positive, and the ratio $b/a=\tan\gamma(0\leqq\gamma\leqq\pi/2)$ is also taken as positive, the upper sign of the right-hand side of (24) corresponds to negative $\sin\delta$, i.e., right-handed rotation, while the lower sign belongs to positive $\sin\delta$, i.e., left-handed rotation.

Thus, in all methods which measure δ and use the relation (24) to determine the ellipticity the sense of rotation results from the sign of the measured phase difference. The wave is right elliptical when $-\pi<\delta<0$, and left elliptical when $0<\delta<\pi$.

Methods with a quarter-wave plate can find the sense of rotation from the orientation of the quarter-wave plate and from the half-shadow azimuth α of the analyzer. If the "fast" axis of the quarter-wave plate is parallel to the major axis of the ellipse and the analyzer is turned from zero position through a negative angle α, the polarization is left elliptical. With crossed position of the two axes and negative rotation the polarization is right elliptical.

With the method of Mac Cullagh-Stokes the sense of rotation is found in a similar way by considering the sign of δ' of the compensator plate and the sign of γ in a compensating position.

When the Szivessy simple half-shade compensator is used to analyze the elliptical wave, the sense of rotation can be found in case the principal direction of vibration of the "faster" ray in the half-shadow plate is known.

IV. INDIRECT METHODS

1. Method of the Soleil Compensator[35]

Jamin used the Babinet compensator to measure the phase difference δ of the elliptical vibration and the ratio γ of the amplitudes of the components of the vibration; from these magnitudes the azimuth ϑ and the ellipticity $\tan\chi$ were calculated.

Since the method of the Soleil compensator is identical to that of the Babinet compensator, but can be used with a half-shadow device, we will discuss the method of the Soleil compensator in conjunction with Szivessy's half-shadow plate.[36]

Determination of the Zero Positions of the Compensator

Compensator and half-shadow plate are placed between crossed nicols in azimuth $\pi/4$ with respect to the vibration direction of the analyzer; by sliding one wedge of the compensator a half-shadow position is found (drum reading p_{01}). If, the corresponding phase difference of the compensator is δ_{01}, then

$$\delta_{01}' = \pm 2n\pi - \eta/2, \quad (n=0, 1, 2, \cdots).$$

Now when the half-shadow plate is rotated through $\pi/2$, a new half-shadow position can be obtained (drum reading p_{02}) and δ_{02}' is the new phase difference of the compensator:

$$\delta_{02}' = \pm 2n\pi + \eta/2.$$

Thus, to a zero position of the compensator $p_0=(p_{01}+p_{02})/2$ corresponds a phase difference

$$(\delta_{01}'+\delta_{02}')/2 = \pm 2n\pi.$$

Measurement of the Phase Difference

Let the elliptically polarized wave be incident upon compensator and half-shadow plate in the zero position p_0. By rotating analyzer and compensator with half-shadow plate simultaneously and by sliding one wedge of the compensator we can find a half-shadow position with weak illumination of the field of view (drum reading p_1, corresponding phase difference of the compensator

$$\delta_1' = -\delta - \eta/2 \pm 2n\pi).$$

In turning the half-shadow plate alone through

[35] Reference 9, p. 961.
[36] G. Szivessy, Zeits. f. Instrumentenk. 46, 454 (1926).

$\pi/2$ another half-shadow position is found (drum reading p_2, corresponding phase difference

$$\delta_2' = -\delta + \eta/2 \pm 2n\pi).$$

The phase difference in question is given by

$$\delta = \frac{p_{01}+p_{02}}{2} - \frac{p_1+p_2}{2}.$$

Measurement of the ratio γ

The compensator is set to the drum reading $p=(p_1+p_2)/2$. In this position the elliptical

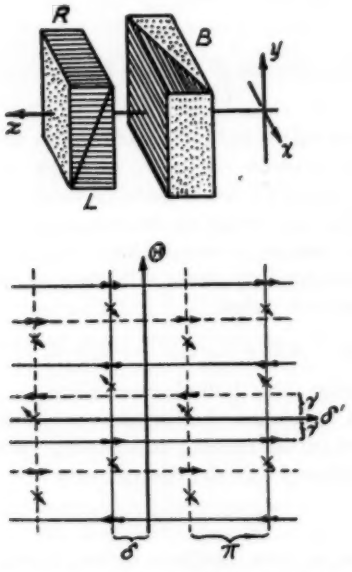


FIG. 11. Voigt's arrangement, and system of zero points.

wave becomes linearly polarized after passing the compensator. To find the vibration direction of this linearly polarized wave the half-shadow plate is replaced by the Nakamura plate and the analyzer is rotated through the angle α_0 to get half-shadow. Thus $\gamma=\alpha_0-\pi/2$. From δ and γ the azimuth ϑ and the ellipticity $\tan\chi$ are calculated by Eq. (24).

2. Voigt's Method[37]

Voigt devised a method for analyzing elliptical polarization which is suitable especially for the ultraviolet region. It consists of a Babinet compensator in conjunction with a rotating quartz wedge plate. This plate is constructed of two exactly equal quartz wedges cut perpendicular to the optic axis, one of dextro- and the other of laevo-quartz, placed together to form a rectangular plate. The edges of this rotating quartz wedge plate are set perpendicular to the edges of the Babinet compensator (Fig. 11).

We discuss the particular case that the wedge plate is introduced between Babinet compensator and analyzer, and that the main directions of vibration of the compensator coincide with the horizontal x-axis and the vertical y-axis. Let the compensator in any section which is parallel to the yz-plane have the phase difference δ'. An elliptical wave propagating in the direction of the z-axis and emerging from the Babinet compensator is given by

$$\mathcal{E}_1 = \begin{pmatrix} 1 \\ \exp(-(\delta+\delta')i) \end{pmatrix} A e^{i\omega t}.$$

When emerging from the wedge plate we have

$$\mathcal{E}_2 = \mathbf{S}(\theta)\mathcal{E}_1,$$

where θ is the rotating angle of the wedge plate; it varies with y. The wave $\mathcal{E}_1$ is linearly polarized when $\delta+\delta'=\pm n\pi$. The vibration direction of this linearly polarized wave emerging from the compensator is given by the azimuth $\pm\gamma$ referred to the x axis, while the corresponding azimuth of the linear vibration emerging from the wedge plate is given by $\phi=\pm\gamma+\theta$. Thus we have the following result. Let P be the point in the xy-plane whereupon the wedges of the compensator as well as that of the wedge plate have the same thickness (Fig. 11). Through P let us draw the δ'-axis parallel to the x axis and the θ-axis parallel to the y axis. The straight lines $\delta'=\pm 2n\pi-\delta$ and $\delta'=\pm(2n+1)\pi-\delta$ are parallel to the θ-axis; the straight lines $\theta=\pm n\pi-\gamma$ and $\theta=\pm n\pi+\gamma$ are parallel to the δ'-axis. The parallels to the θ-axis give the loci of linear polarization. The intersections of the two systems of parallels are the points for $\phi=\pm n\pi$; their vibration directions are drawn by short arrows. The points with azimuths $\phi=\pm(2n+1)\pi/2$ lie between them (not drawn in Fig. 11). The wave emerging from the analyzer is extinguished at the points where its vibration direction is perpendicular to that of the analyzer. When the analyzer azimuth is $\pi/4$, points marked by crossed arrows are dark. In rotating the analyzer these "zero"-points are displaced along the

[37] W. Voigt, Physik. Zeits. 2, 303 (1901); see reference 9, p. 963.

straight line parallel to the θ-axis. On the other hand, when the analyzer remains fixed and the values of δ and γ are varied, the zero-points are shifted.

The experimental proof of Voigt's method was carried out by Minor[38] as follows. Let first a linearly polarized wave of the same frequency as that of the elliptical vibration in question fall in azimuth $\pi/4$ on the Voigt arrangement. In taking a photograph the zero positions are fixed. Let then the elliptical wave fall on the same arrangement and again a photograph is taken. The displacement of the zero-points on the second picture with respect to the first is measured and gives the values of δ and γ. Azimuth ϑ and ellipticity $\tan\chi$ can be calculated by means of Eqs. (24).

[38] R. S. Minor, Ann. d. Physik **10**, 581 (1903).

JOURNAL OF THE OPTICAL SOCIETY OF AMERICA VOLUME 39, NUMBER 2 FEBRUARY, 1949

Scattered and Rescattered Sunlight

A. HAMMAD
Department of Mathematics, Faculty of Science, Cairo, Egypt

(Received October 11, 1948)

In this paper the functions involved in the emission and reception of secondary scattered (rescattered) sunlight are discussed. *Index-sum* approximations of the form

$$\sum_{r=1}^{r=3} B_{nr}\exp(-\beta_{nr}x)$$

are obtained for the integrals $Ej_n(x)$, defined by

$$Ej_n(x)=\int_1^{\infty}\exp(-xt)t^{-n}dt.$$

The factors β_{nr} are determined numerically by Prony's method of interpolation by exponentials, and the coefficients B_{nr} by the method of least squares and Seidel's method of iteration. The functions $Ej_n(x)$ occur as factors in the integrands of the integrals

$$L_n(c,z,\bar{m})=\int_0^1\exp(-Z\bar{m}')Ej_n(c\bar{m}\sim c\bar{m}')d\bar{m}'$$

involved in the emission of secondary scattered light. By means of the index-sum approximations for the integrals $Ej_n(x)$, approximate formulas for the functions L_n are obtained. Numerical tables are given for Ej_n and L_n, showing the high degree of accuracy in our approximations. Likewise, the functions L_n occur as factors in the integrands of the integrals Lu_n and Ld_n involved in the reception of secondary scattered light. These integrals are of the type

$$\int_0^X\exp(bx)L_n(c,z,x)dx.$$

Hence, approximate formulas for such integrals can be obtained in which the degree of accuracy is higher than that obtained for the functions L_n.

According to the approximate formulas given in this paper, an estimate can be made for secondary scattered sunlight in which the error is less than one percent of the exact value.

The formulas apply to all solar altitudes and to a wide range of wave-lengths in the solar spectrum, in contrast with previously obtained approximate formulas which apply to weakly attenuated radiations and high solar altitudes.

I. INTRODUCTION

IN a previous paper[1] (here referred to as Paper I) an exact solution was given of the problem of the passage of sunlight through a plane-stratified atmosphere of uniform composition, as far as primary and secondary scattering are concerned, taking account of Rayleigh's law of scattering and neglecting the polarization of radiation by scattering.[2] The formulas for the reception of primary scattered light are quite simple and easy to evaluate, but those for secondary are so complicated and elaborate to evaluate that the need is urged for much simpler expressions whose numerical values agree with those of the exact formulas within assigned close limits. Attempts towards that end were made by L. Silberstein,[3] and more recently by the present writer,[4] but both considered a very special case, namely, the case when the sun is overhead and the radiation is weakly attenuated. Their formulas are obviously inapplicable to moderate or low solar altitudes, particularly when the attenuation coefficient is not small. The results of a

[1] A. Hammad and S. Chapman, Phil. Mag. **28**, Ser. 7, 99 (1939).

[2] On this account, the formulas for secondary scattered light need modification if the polarization of radiation by scattering is taken into consideration. This is the subject of a paper to appear soon in The Astrophysical Journal.

[3] L. Silberstein, J. Opt. Soc. Am. **33**, 526 (1943).

[4] A. Hammad, J. Opt. Soc. Am. **38**, 611 (1948).

recent investigation published elsewhere,[5] concerning the relative magnitudes of the different orders of scattering, show, however, that the cases considered by those two authors are really of very little importance, as the contribution of secondary scattering in such cases is negligibly small compared to that of the primary. On the other hand, in the case of radiations for which the attenuation coefficient of the atmosphere is not small, the contribution of secondary scattering is comparable to that of the primary, particularly for low solar altitudes. Such cases are thus of special interest, and simple approximate formulas for secondary scattering are really very desiable in these cases.

In the present paper simple approximate expressions for secondary scattered light are derived, which apply to all solar altitudes, and to a wide range of values of the attenuation coefficient, corresponding to a considerable part of the solar spectrum. Other approximate expressions have been given elsewhere,[6] but the expressions obtained here have the advantage of more generality.

The symbols c, z, $\bar{m}$, Z which will occur in some of our formulas have the same meaning as in Paper I: c denotes the coefficient of total extinction of the atmosphere for vertically incident light (of wave-length λ), z the zenith angle of the sun, $\bar{m}$ the fraction of the total mass of the atmosphere that lies above any point, and $Z=c\sec z$.

In this paper we shall consider radiations for which c does not exceed 1; these correspond to a wide range of wave-lengths in the solar spectrum ($\lambda \gtrsim 3000$A).

II. THE EMISSION OF SECONDARY SCATTERED SUNLIGHT

The theory of emission of secondary scattered sunlight introduces the integrals

$$L_n(c,z,\bar{m})=\int_0^1 \exp(-Z\bar{m}')Ej_n(c\bar{m}\sim c\bar{m}')d\bar{m}',$$

where

$$Ej_n(x)=\int_1^\infty \exp(-xt)t^{-n}dt, \quad \cdots, \qquad (1)$$

for $n=1$, 3, and 5 (cf. Section 13, Paper I).

The functions L_n are specified and discussed in Section 17 of Paper I, but the expressions for them are rather complicated. Our main object in the present paper is to replace these functions by simpler approximate expressions whose numerical values agree with those of the exact expressions within close limits. It is easy to see that in the integrals L_n the integrand is the product of two factors—the exponential term $\exp(-Z\bar{m}')$ and the function $Ej_n(c\bar{m}\sim c\bar{m}')$. It seems natural, therefore, to seek approximate expressions for $Ej_n(x)$ in the form of sum of exponentials.

III. APPROXIMATE FORMS FOR THE INTEGRALS $Ej_n(x)$

The integral $Ej_1(x)$ is closely related to the well known exponential integral $Ei(-x)$, which was calculated and discussed by Glaisher, Dale, and others; in fact, $Ej_1(x)=-Ei(-x)$. Extensive tables* for this integral over a wide range of x are now available.

By means of recurrence relations (given in Section 15 of Paper I), $Ej_n(x)$ for any positive integer n can be calculated. Gold[7] gave tables of $Ej_n(x)$, from $n=1$ to $n=3$, to five decimal places, from $x=0$ to $x=6$. the interval being 0.01 from $x=0$ to $x=0.1$, and 0.05 from $x=0.1$ to $x=6$. More recently we have had occasion[5] to publish tables of $Ej_n(x)$ from $n=1$ to $n=5$, to six decimal places, at interval 0.01, from $x=0$ to $x=1$.

From the definition of $Ej_n(x)$ it may be readily seen that it decreases as either x or n increases. Further, $Ej_n(x)$ is finite for all (positive) values of x if $n>1$; for $n=1$, however, $Ej_1(x)\to\infty$ like $\log x$ as $x\to 0$ (cf. Section 15, Paper I).

If in Eq. (1) the variable t is changed into v ($v=1/t$), the integral (1) becomes

$$Ej_n(x)=\int_0^1 \exp(-x/v)v^{n-2}dv, \quad \cdots, \qquad (2)$$

which shows that the integrand as well as the interval are finite for $n>1$.

Consider now the case $n>1$; if we divide the range of integration in (2) into three parts, we get

$$Ej_n(x)=\left(\int_0^{v_1}+\int_{v_1}^{v_2}+\int_{v_2}^1\right)\times\exp(-x/v)v^{n-2}dv. \qquad (3)$$

[5] A. Hammad, Phil. Mag. **38**, Ser. 7, 515 (1947).
[6] A. Hammad and S. M. Hassanein, Phil. Mag. (to be published).

* For the range 0...1, see reference 5. For the range 0...15, see British Association *Mathematical Tables* (London 1931) Vol. 1, also *Tables of Sine, Cosine and Exponential Integrals* (Work Project Administration, National Bureau of Standards, 1940). For the range 15...50, see, Coulson and Duncanson, Phil. Mag. **33**, Ser. 7, 754 (1942).
[7] E. Gold, Proc. Roy. Soc. **82**, Ser. A, 62 (1909).

TABLE I. Values of B_{nr}, β_{nr}, $\exp(-0.2\beta_{nr})$.

	$n=3$			$n=5$		
	$r=1$	$r=2$	$r=3$	$r=1$	$r=2$	$r=3$
B_{nr}	=0.359687	0.124907	0.015417	0.184253	0.062977	0.002769
β_{nr}	=1.251289	2.914563	9.859218	1.118589	1.818283	4.551294
$\exp(-0.2\beta_{nr})$	=0.778600	0.558270	0.139200	0.799540	0.695130	0.402420

By the mean value theorem,[8] each of the three partial integrals in (3) can be replaced by a quantity of the form

$$B\exp(-\beta x),$$

so that

$$Ej_n(x)=\sum_{r=1}^{r=3} B_{nr}\exp(-\beta_{nr}x). \qquad (4)$$

It is convenient to call (4) *an index-sum*, and B_{nr}, β_{nr} its index coefficients and index factors, respectively.

The index coefficients and the index factors will be determined in this paper by numerical methods. For the index factors β_{nr} we shall adopt Prony's method of interpolation by exponentials.[9] Given the values of $Ej_n(x)$ for a set of values of x, say, x_0, x_0+w, x_0+2w, $\cdots$, x_0+sw (such that the difference between any two successive values of x is constant), and denoting the corresponding values of $Ej_n(x)$ by E_{n0}, E_{n1}, E_{n2}, $\cdots$, E_{ns}, respectively, Prony's method determines the quantities

$$\exp(-\beta_{n1}w),\quad \exp(-\beta_{n2}w),\quad \exp(-\beta_{n3}w)$$

as the roots of the cubic equation

$$t^3+Pt^2+Qt+S=0,$$

where P, Q, S satisfy the set of linear equations

$$\begin{aligned}
&E_{n3}+PE_{n2}+QE_{n1}+SE_{n0}=0,\\
&E_{n4}+PE_{n3}+QE_{n2}+SE_{n1}=0,\\
&E_{n5}+\cdots,\\
&\cdots\\
&\cdots\\
&E_{ns}+PE_{n,s-1}+QE_{n,s-2}+SE_{n,s-3}=0.
\end{aligned}$$

By the method of least squares, or otherwise, P, Q, S can be determined from the above set of equations, and subsequently the roots of the above cubic equation. Having thus determined the index factors, the index coefficients B_{nr} can be determined by the method of least squares or Seidel's method of iteration.**

In this way we have determined β_{nr} and B_{nr} for $n=3$ and $n=5$, taking values of $Ej_n(x)$ at $x=0$, 0.2, 0.4, 0.6, 0.8, and 1. In Table I the values of B_{nr}, β_{nr} are given, together with the values of $\exp(-0.2\beta_{nr})$, for the purpose of simplifying the computation.

It must be remembered, however, that the approximate forms for Ej_3 and Ej_5, thus obtained, are only valid for that range of x ($0\leqslant x\leqslant 1$) defined by the points which were selected for the determination of the index factors and index coefficients. Outside this range these approximations are not trustworthy.

Using these values of B_{nr} and β_{nr}, we have calculated some values of $Ej_3(x)$ and $Ej_5(x)$; these are compared in Table II with the exact values (taken from a recently published paper[5]).

In the case of $Ej_5(x)$ it is clear from this table that the differences are negligibly small, there being no error in the fifth decimal place. In the case of $Ej_3(x)$ the errors are also small, though they are slightly higher in the interval $0\cdots 0.2$.

Index-sum approximations of five terms, obtained by a different method, were given by Mian and Chapman[8] for the integrals $Ej_n(x)$ for $n>1$. For $Ej_1(x)$, however, no index-sum approxi-

TABLE II. Exact and approximate values of $Ej_3(x)$, $Ej_5(x)$.

x	=0	0.2	0.4	0.6	0.8	1.0
$Ej_3(x)$ Exact	=0.50000	0.35195	0.25729	0.19155	0.14432	0.10969
$Ej_3(x)$ Approx.	=0.50001	0.35193	0.25728	0.19155	0.14432	0.10970
$Ej_5(x)$ Exact	=0.25000	0.19221	0.14867	0.11551	0.09007	0.07045
$Ej_5(x)$ Approx.	=0.25000	0.19221	0.14867	0.11551	0.09007	0.07045

[8] The credit of this argument is due Dr. M. N. Mian and Professor S. Chapman [Phil. Mag. 33, Ser. 7, 115 (1942)].
[9] E. T. Whittaker and G. Robinson, *The Calculus of Observation* (Blackie and Sons, Ltd., London, 1924), fourth edition (1946), p. 369.
** See reference 9, p. 255.

TABLE III. Exact and approximate values of $Ej_1(x)$.

x		=0.01	0.2	0.4	0.6	0.8	1.0
$Ej_1(x)$	Exact	=4.03793	1.22265	0.70238	0.45438	0.31060	0.21938
	Approx.	=4.03793	1.22250	0.70269	0.45642	0.31297	0.21938

mation has been obtained, probably because of the behavior of this function near $x=0$. But if a small interval, say $0\cdots0.01$ (including $x=0$), is excluded, it will be possible to represent $Ej_1(x)$ by an index-sum of three terms, valid in the range $0.01\cdots1$. Taking values of $Ej_1(x)$ at $x=0.01, 0.11, 0.21, 0.31, 0.41, \cdots, 1.01$, we have obtained by Prony's method and the method of least squares the following values for B_{1r}, β_{1r} and $\exp(-0.1\beta_{1r})$.

$r=$	1	2	3
$B_{1r}=$	2.11752	1.37416	1.20141
$\beta_{1r}=$	29.83560	6.65038	1.70864
$\exp(-0.1\beta_{1r})=$	0.050612	0.514255	0.842943

Using these values of B_{1r}, β_{1r} we have calculated some values of $Ej_1(x)$; these are given in Table III, together with the exact values for the sake of comparison.

It is clear from this table that the error in the case of Ej_1 is also very small, being less than one percent of the exact value in all the cases.

IV. APPROXIMATE FORMS FOR $L_n(c, z, \bar{m})$

We shall now apply the approximations given in the previous article to get approximate forms for the functions L_n involved in the emission of secondary scattered sunlight.

As mentioned in Section II above, the functions L_n are defined by the integral

$$L_n(c, z, \bar{m}) = \int_0^1 \exp(-Z\bar{m}')Ej_n(c\bar{m}\sim c\bar{m}')d\bar{m}';$$

by a change of variable this becomes

$$c\exp(Z\bar{m})L_n = \int_0^{c\bar{m}} \exp(\sec z\, x)Ej_n(x)dx + \int_0^{cm} \exp(-\sec z\, x)Ej_n(x)dx, \quad (1)$$

where $m=1-\bar{m}$.

Substituting the index-sum approximations for $Ej_n(x)$, we obtain the following approximate form for L_n, valid for $n=3$ and $n=5$:

$$L_n(c, z, \bar{m}) = \sum_{r=1}^{r=3} B_{nr}[\{\exp(-c\beta_{nr}\bar{m}) - \exp(-Z\bar{m})\}/(Z-c\beta_{nr}) + \{\exp(-Z\bar{m}) - \exp(-Z-c\beta_{nr}m)\}/(Z+c\beta_{nr})]. \quad (2)$$

A similar approximate form for L_1 cannot be assumed without justification, as the index-sum approximation of Ej_1, given in the previous article, is valid only in the range $0.01\cdots1$, and not in the whole ranges of integration in (1). We have, therefore, to justify its extension to the small range of integration $0\cdots0.01$. It is clear from (1) that in this small range we have, in general, two integrals of the type:

$$\int_0^{0.01} \exp(bx)Ej_1(x)dx, \quad (3)$$

where b stands for $\pm\sec z$. The exact value of this integral lies between the two limits

$$\{\exp(0.01b), 1\}\int_0^{0.01} Ej_1(x)dx,$$

i.e., between

$$\{\exp(0.01b), 1\}\times\{\tfrac{1}{2}-Ej_2(0.01)\},$$

or

$$\{\exp(0.01b), 1\}\times 0.05033.$$

Asuming the approximate form of $Ej_1(x)$ to apply to the range $0\cdots0.01$, we get for the integral (3) the approximate expression

$$\sum_{r=1}^{r=3} B_{1r}\{\exp(0.01(b-\beta_{1r}))-1\}/(b-\beta_{1r}).$$

Taking $\sec z=1$ and choosing the positive sign for b, we have calculated the upper and lower limits of the integral (3), and also its approximate value; these are, respectively, 0.0508, 0.0503, and 0.0437. Thus, for this particular value of z, the difference between the exact and approximate value is roughly 14 percent of the exact value. Nearly the same value for the difference is obtained for other values of z. Such magnitudes for

the difference do not affect appreciably the value of L_1, as the range of integration in (3) is small compared with those in (1). This conclusion is endorsed numerically by Table IV, where the exact and approximate values of L_1, L_3, and L_5 are given for some values of c, z, and $\bar{m}$, the exact values being given *in italics*.

It can be seen from this table that the differences between the exact and approximate values of L_1 are really very small, being less than one percent of the exact value. Higher accuracy may be expected in the cases of L_3 and L_5; Table IV shows that this is so, the differences between the exact and approximate values of L_3 and L_5 being negligibly small.

Thus, on the bases of the approximate formulas for $Ej_n(x)$, given in the previous article, an estimate can be made of the intensity of emission of secondary scattered sunlight, in which the error is less than one percent.

TABLE IV. Exact and approximate values of $L_n(c, z, \bar{m})$ (*exact values in italics*).

Values of L_1

	$c=0.1$			$c=1$	
sec z	$\bar{m}=0$	$\bar{m}=0.4$	$\bar{m}=1$	$\bar{m}=0$	$\bar{m}=1$
1	2.66801	3.24146	2.62655	0.66219	0.43278
	2.66349	*3.26745*	*2.61746*	*0.66134*	*0.43173*
6	2.19164	2.58776	1.99599	0.32349	0.05275
	2.19873	*2.61145*	*1.98154*	*0.32425*	*0.05250*

Values of L_3

	$c=0.1$		$c=1$	
sec z	$\bar{m}=0$	$\bar{m}=1$	$\bar{m}=0$	$\bar{m}=1$
1	0.43484	0.43351	0.17564	0.13810
	0.43459	*0.43327*	*0.17562*	*0.13810*
6	0.34621	0.33998	0.06455	0.02359
	0.34601	*0.33978*	*0.06453*	*0.02359*

Values of L_5

	$c=0.1$		$c=1$	
sec z	$\bar{m}=0$	$\bar{m}=1$	$\bar{m}=0$	$\bar{m}=1$
1	0.22305	0.22256	0.09803	0.08002
	0.22305	*0.22257*	*0.09803*	*0.08002*
6	0.17722	0.17491	0.03418	0.01466
	0.17722	*0.17490*	*0.03418*	*0.01466*

V. THE RECEPTION OF SECONDARY SCATTERED LIGHT

The reception of secondary scattered sunlight introduces further integrals Lu_n and Ld_n (cf. Section 14, Paper I), which are of the type

$$\int_0^x \exp(bx) L_n(c, z, x)dx.$$

It is clear that with any approximate form for L_n, the error in an integral of the above type will not exceed the maximum error in the corresponding L_n. Thus, on the bases of the approximate forms of $Ej_n(x)$, given in Section 3, the errors in Lu_3, Ld_3, Lu_5, Ld_5 will be negligible, while for Lu_1 and Ld_1 the error will not exceed one percent of the exact value. With this in view, we conclude that our approximation gives numerical values for secondary scattered sunlight in very close agreement with the exact values.

The Optical Theory of the Spectral Sensitivity of a Caesium Oxide Photo-Cathode

YOSHIHIRO SAYAMA
Electronic Engineering Research Laboratory, Tokyo Shibaura Electric Company, Ltd., Tokyo, Japan
(Received November 19, 1947)

Silver which is dispersed as small particles in the intermediate layer of Cs_2O photo-cathode has the characteristic optical constants of a thin film and hence has an absorption band in visible range. On the other hand, the spectral sensitivity of the cathode shows a remarkable maximum in visible range. We assume that the selective maximum is due to the volume photo-electric effect corresponding to the foregoing absorption band of silver thin film. We have calculated the light absorption in the surface layer of a silver thin film lying on a glass plate and obtained the spectral sensitivity which is qualitatively in agreement with the experimental one.

INTRODUCTION

IT is well known that the spectral sensitivity of Cs_2O photo-cathode has a remarkable maximum at 6000 to 8000A, and its threshold lies at about 12000A. Such a selective maximum of the spectral sensitivity is also observed for all metal surfaces covered with alkali metal films. Ives[1] interpreted successfully this selective effect of thin films of alkali metal calculating the electric vector of light on the surface and assuming that the number of photoelectrons was proportional to the absorption of light in the surface layer of the cathode.

In this article the author tries to show that the selective effect of a Cs_2O photo-cathode can also be interpreted by such an optical factor.

STRUCTURE OF Cs_2O PHOTO-CATHODE

Ives[2] investigated the spectral sensitivity of this cathode by the above method and concluded that the selective effect in this case could not be interpreted by the optical factor. He employed a simplified model—a platinum base metal covered with an intermediate layer of quartz, whose surface is covered with a thin photosensitive film of caesium—as the structure of the Cs_2O photo-cathode. Such a model is, however, far from the real structure of a Cs_2O photo-cathode. Therefore, we want to start with a discussion of the structure of a Cs_2O photo-cathode.

A Cs_2O photo-cathode is usually prepared by the following method. A silver oxide film of about fifty molecular layers is prepared on the surface of a thick silver film by the discharge in the oxygen gas. The caesium vapor is introduced to the surface of the oxide layer. Then the silver oxide is reduced and the fine silver particles are dispersed among the layer of the cassium oxide. This cathode can be further sensitised by the deposition of a small quantity of silver upon it.[3] The spectral sensitivity of this cathode is shown in Fig. 1[4] (curve 1). In this type of cathode the boundary plane between the base of the silver metal and the Cs_2O layer is not as smooth because of the oxidation and the reduction, so it is not adequate to our optical calculation.

Another type of Cs_2O photo-cathode is the semitransparent one; which is prepared by the following method. The silver film of a few mμ thickness is deposited on the surface of a glass wall, and then it is thoroughly oxidized. The subsequent procedure is similar to the first type of cathode. In this case the Cs_2O layer lies directly on the glass surface and the layer contains the silver particles. The spectral sensitivity of this type is shown in Fig. 1[5] (curve 2). We shall try to calculate the optical factor for this type, as the structure of this type is simpler than the first. But there is no essential difference in the mechanism of the electron emission between these types of cathodes.

According to Asao,[6] the Cs_2O layer which contains no silver particles shows no significant absorption in the visible spectral range. On the other hand, the silver film evaporated on this layer for the purpose of the sensitization shows a transmission and reflection characteristic to the thin film of silver and not to the thick one. The spectral sensitivity of this cathode is curve 3

[1] H. E. Ives, Phys. Rev. **38**, 1209 (1931). H. E. Ives and H. B. Briggs, Phys. Rev. **38**, 1477 (1931); **40**, 802 (1932); J. Opt. Soc. Am. **26**, 247 (1936); **28**, 330 (1938).
[2] H. E. Ives and A. R. Olpin, J. Opt. Soc. Am. **24**, 198 (1934).
[3] S. Asao, Proc. Phys. Math. Soc. Jap. **22**, 448 (1940).
[4] Y. Sayama, Toshiba Kenkyu Jiho **19**, 219 (1944).
[5] H. Miyazawa, Toshiba Review **2**, 49 (1947).
[6] S. Asao, Proc. Phys. Math. Soc. (Japan) **22**, 448 (1940).

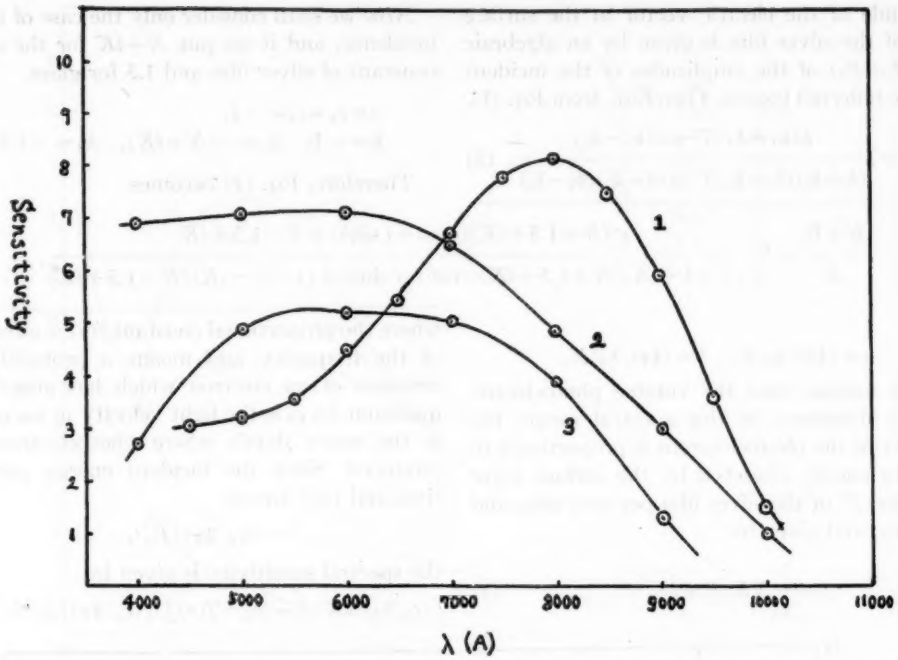


FIG. 1. Spectral sensitivity of Cs_2O photo-cathodes of various types. Curve 1: [Ag]−Cs_2O, Ag, Cs−Ag; curve 2: [Glass]−Cs_2O, Ag, Cs−Ag; curve 3: [Quartz]−Cs_2O, Cs−Ag.

(Fig. 1). According to Miyazawa's exhaustive study[7] of the semitransparent Cs_2O cathode, the reduced silver particles in the Cs_2O layer play a similar role, photoelectrically or optically, as the silver film deposited for the sensitization. Thus we are inclined to take up the following simplified model for this photo-cathode. On the glass surface lies a silver thin film of about 100A thickness which has the optical constants not of bulk state but of thin state, and whose photoelectric threshold wave-length lies at about 12000A because of the very thin covering film of Cs_2O. We completely neglect the optical behavior of Cs_2O (Fig. 2).

CALCULATION

When the incident light whose electric vector E lying perpendicular to the incident plane is thrown upon this silver film (Fig. 2), the complex amplitude of the reflected beam E_2 is given by the formula:[8]

$$\frac{E_2}{E}=\frac{(k-k_1)(k_1+k_4)X^2+(k+k_1)(k_1-k_4)}{(k+k_1)(k_1+k_4)X^2+(k-k_1)(k_1-k_4)} \quad (1)$$

where k's and X are

$$k=(N+iK)c,$$
$$X=\exp\{-(2\pi i/\lambda_0)[(N+iK)t]\}.$$

$(N+iK)$ is the optical constant of each medium, c is the direction cosine of the direction of each beam with respect to z-axis which is taken perpendicular to the boundary plane toward the vacuum side, t is the thickness of the silver film, λ_0 is the wave-length in vacuum, and the suffixes mean the quantities concerning the beam with the same suffixes in Fig. 2 or concerning the medium in which these beams are passing. The

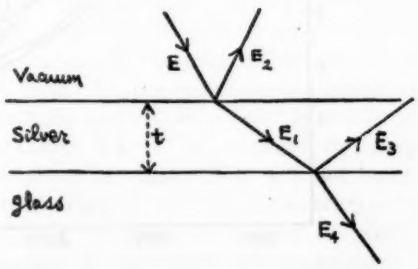


FIG. 2. Simplified model of semitransparent Cs_2O photo-cathode.

[7] H. Miyazawa, Toshiba Review 2, 49 (1947).
[8] T. C. Fry, J. Opt. Soc. Am. 22, 307 (1932).

amplitude of the electric vector in the surface layer of the silver film is given by an algebraic sum $(E+E_2)$ of the amplitudes of the incident and the reflected beams. Therefore, from Eq. (1),

$$\frac{E+E_2}{E}=2\frac{k(k_1+k_4)X^2+k(k_1-k_4)}{(k+k_1)(k_1+k_4)X^2+(k-k_1)(k_1-k_4)}. \quad (2)$$

Now we shall consider only the case of normal incidence, and if we put $N+iK$ for the optical constant of silver film and 1.5 for glass,

$$c=c_1=c_4=-1,$$
$$k=-1, \quad k_1=-(N+iK), \quad k_4=-1.5.$$

Therefore, Eq. (2) becomes

$$\frac{E+E_2}{E}=2\frac{e^a(N+1.5+iK)(\cos b-i\sin b)+N-1.5+iK}{e^a(N+1+iK)(N+1.5+iK)(\cos b-i\sin b)+(1-N-iK)(N-1.5+iK)}, \quad (3)$$

where

$$a=(4\pi t/\lambda_0)K, \quad b=(4\pi t/\lambda_0)N.$$

If we assume that the volume photoelectric effect is dominant in this spectral range, the intensity of the photoelectrons is proportional to the light energy absorbed by the surface layer (thickness t') of the silver film per unit time and unit area, and given by

$$f(\nu)\frac{c_0}{\lambda_0}NK|E+E_2|^2t', \quad (4)$$

where the proportional constant $f(\nu)$ is a function of the frequency and means a probability of emission of an electron which has absorbed a quantum $h\nu$. c_0 is the light velocity in vacuum. t' is the mean depth where photoelectrons are produced. Since the incident energy per unit time and unit area is

$$(c_0/4\pi)|E|^2,$$

the spectral sensitivity is given by

$$[(c_0/\lambda_0)NK|E+E_2|^2t'f(\nu)]/[(c_0/4\pi)|E|^2h\nu]. \quad (5)$$

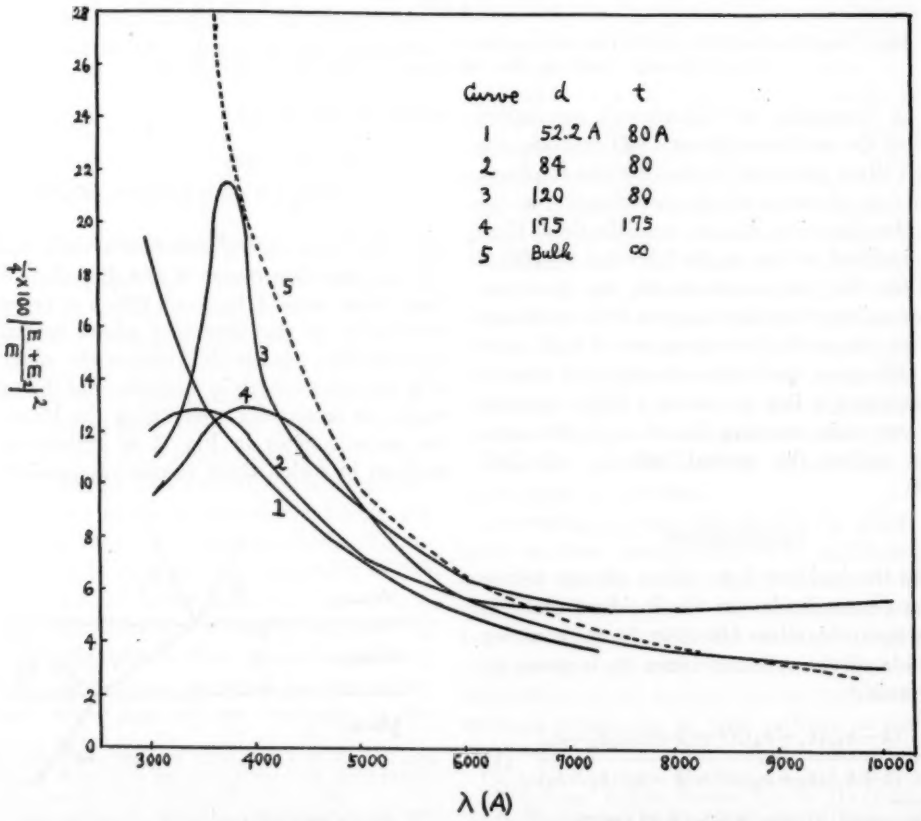


FIG. 3. Light energy in the surface layer of silver film.

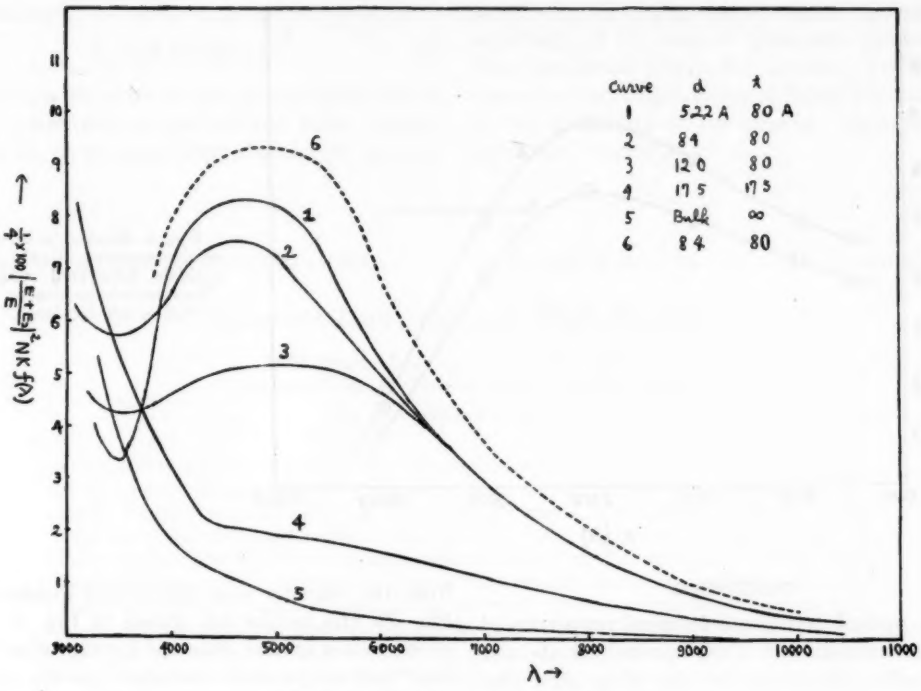


FIG. 4. Spectral sensitivity (calculated).

At present we have no information on the explicit form of the function $f(\nu)$. But if we assume that the probability of absorption of light is independent of the initial value of the electronic wave number, then

$$f(\nu) \propto (\nu-\nu_c)^2,$$

where ν_c is the threshold frequency. After all, the spectral sensitivity S is given by

$$S=NK(\nu-\nu_c)^2|(E+E_2)/E|^2 \qquad (6)$$

where we omit the factors independent of frequency. The last factor is given by Eq. (3).

COMPARISON WITH EXPERIMENT

It is well known that silver has abnormal optical constants in the state of thin film. It has a strikingly large index of refraction. We use Goos-Murmann's values[9] of N and K and put in Eqs. (3) and (6). The results are shown in Figs. 3 and 4, respectively. The threshold frequency was put $\nu_c=2.43\times10^{14}$ sec.$^{-1}$($\lambda_c=12340$A). In these figures, d means the thickness at which Goos measured the optical constants and t is the value substituted in (3) and (6). The values do not always coincide, since the grain size of silver particles in the photo-cathode may change owing to the oxidation, reduction, and heat treatment. For N and K of bulk silver we use Minor's values.[10] Figure 3 shows that if the characteristic optical constants of thin film ($d\leq120$A) are used, there happens the selective maximum of spectral sensitivity at about 5000A to 6000A. This is rather shorter as compared with the experimental value (Fig. 1). But considering the bold simplification assumed at first, it can be said that the calculated spectral sensitivity is at least qualitatively in agreement with the experimental one. Since $|E+E_2|^2/|E|^2$ values show no preferable enhancement in the selective maximum for thin film (Fig. 2), the selective maximum of the calculated spectral sensitivity is mainly due to the factor NK in (3), namely the degree of absorption of light. The spectral sensitivity corresponding to the optical constants of thicker film (Fig. 3 curve 4.5) shows no pronounced maximum in conformity with experience.

[9] F. Goos, Zeits. f. Physik **100**, 95 (1936); H. Murmann, Zeits. f. Physik **101**, 643 (1936).

[10] Minor, Ann. d. Physik **10**, 581 (1903).

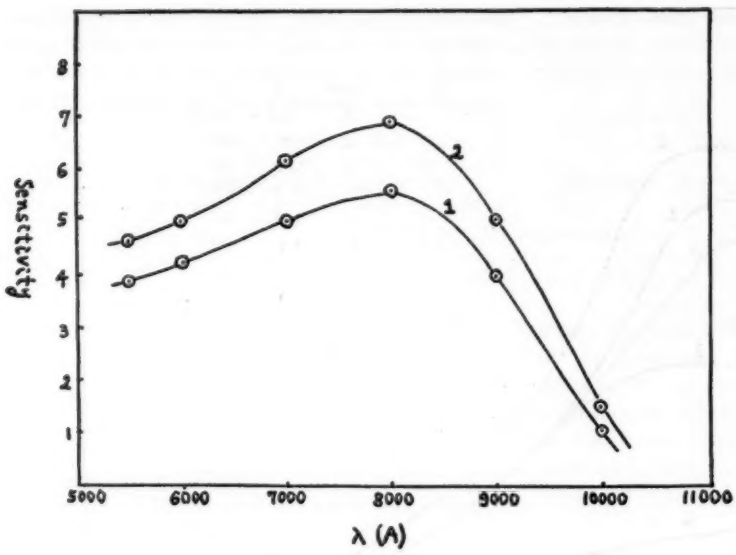


FIG. 5. Spectral sensitivity of semitransparent Cs_2O Photo-cathode (observed). Curve 1: "Vacuum side incidence;" curve 2: "Glass side incidence."

DISCUSSION

The optical and photoelectrical properties of the semitransparent Cs_2O photo-cathode are principally determined by the silver thin film consisting of the reduced silver particles and the sensitizing silver particles, whose surfaces are covered with thin Cs_2O layer and consequently have low work function nearly 1 electron volt. The photoelectrons originate from these silver particles and the process is considered to be the volume photoelectric effect.

In texts[11] the light absorption by metal in visible spectral range is treated usually as the classical damped oscillation of free electrons. But this is insufficient for the theoretical treatment of the volume photoelectric effect in this spectral range. Muto[12] carryed out the quantum-mechanical calculation of this process. But this was not yet compared with any experiment.

The serious weak point of Ives' investigations and also for the author's article exists in this point. It is desirable that the quantum-mechanical treatments of this process are developed.

APPENDIX

According to Hirai,[13] the higher sensitivity of the semitransparent Cs_2O photo-cathode is obtained when the incident light is thrown from the base side (upward incident in Fig. 2) rather than from the vacuum side (downward incidents in Fig. 2). His results are shown in Fig. 5. This phenomenon is also observed for the other thin film[14] and can be easily deduced from the optical factor as follows.

Now the case of "base side incidence" is shown schematically in Fig. 6. If we consider the case of normal incidence, the amplitude of E vector in the surface layer of silver is equal to the amplitude of the beam 6 and given by[8]

$$E_6/E_3=4k_3k_5X/[(k_3+k_5)(k_5+k_6)X^2+(k_3-k_5)(k_5-k_6)], \quad (7)$$

where E_3 is the amplitude of beam 3 (incident beam in glass), X and k's have same meanings as (1) and we neglect the back reflexion of beam 4 at the glass-air boundary plane. The amplitude E_3

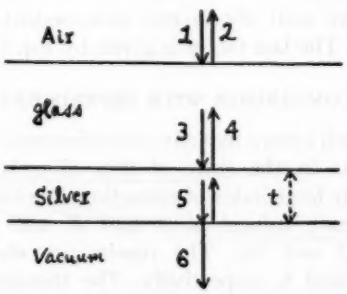


FIG. 6. Schematic diagram of the case of "glass side incidence."

[11] Cf. H. Fröhlich, *Elektronen Theorie der Metalle* §8.
[12] T. Muto, Sci. Pap. Inst. Phys. Chem. Research 27, 179 (1935).
[13] S. Hirai, Mazda Kenkyu Jiho 16, 10 (1941).
[14] P. Görlich, Zeits. f. Physik 101, 335 (1936); A. L. Hughes and L. A. DuBridge, *Photoelectric Phenomena* (1932), p. 122.

is given by

$$E_3=[2/(N_0+1)]E_1, \quad (8)$$

where N_0 is the index of refraction of glass and E_1 is the amplitude of the incident beam. Substituting E_6/E_1 obtained from (7) and (8) in place of $(E+E_2)/E$ in Eq. (6), we obtain the spectral sensitivity of the case of "glass side incidence." The calculated curve for $d=84$A, $t=80$A is shown in Fig. 4 (curve 6) and forms the contrast to the sensitivity of the case of "vacuum side incidence" (curve 2).

JOURNAL OF THE OPTICAL SOCIETY OF AMERICA VOLUME 39, NUMBER 2 FEBRUARY, 1949

Absolute Thresholds and Night Myopia

J. M. Otero, L. Plaza, and F. Salaverri
Institute of Optics, "Daza de Valdés," Madrid, Spain

(Received September 28, 1948)

The first measurements of night myopia showed a clear decrease in the minimum quantity of light needed to reach a certain visual acuity when the night myopia was compensated. The results indicated a possible decrease in absolute thresholds. In the present paper we study, with four observers using binocular vision and white light, the influence of night myopia on these thresholds for target areas ranging from 30 seconds to 10.5 minutes and retinal zones from the fovea up to 15° eccentricity. A statistical technique secures very uniform results. We find unquestionably that the values of the threshold decrease substantially in every case when night myopia is compensated. The average diminution is 50 percent. On the other hand, the abnormal light distribution on the retina caused by night myopia, when compensated, may modify the laws relating absolute thresholds with stimulus area. We have studied these possible changes in the generally admitted range of validity of Riccó's law without finding any noticeable variation.

INTRODUCTION

SINCE the first quantitative measurements of night myopia,[1] it could be verified that on correcting this anomaly the brightness needed to achieve a fixed *minimum cognoscible* dropped considerably. Subsequently it was concluded that the visual acuity[2–4] obtainable at fixed brightness increases considerably once night myopia has been compensated. Ronchi[5] in 1943 observed that many more stars could be seen when this ametropia is compensated by suitable spectacles than when observing with the naked eye. All these facts indicate that the absolute threshold must be influenced by such ametropia. This influence has to be further stressed because of the existence of another phenomenon discovered by Durán in 1943,[6] *viz.*, nocturnal presbyopia, which was also experimentally confirmed by Cabello and Ivanoff in 1944 and 1946. It consists of a progressive fall of accommodating power as the brightness decreases, arriving to a complete presbyopia near the threshold.

Night presbyopia and myopia must leave the refraction of one eye fixed in darkness, with its far point at a close distance, thus bringing about blurring of the images of objects located at long distances, and producing from an object point not an image of the same kind distorted exclusively by diffraction and lens aberration, but a much larger spurious disk. Therefore we may foresee a descent in the absolute threshold corresponding to that for a point or near-point source when night myopia is compensated. This decrease will lose importance progressively as the object enlarges and the spurious spreading of the retinal image diminishes relative to its whole area.

The problem has very interesting theoretical and practical aspects. Measurements of the absolute thresholds of near-point sources—i.e., those in which Riccó's law is considered valid—have been the object of recent investigations[7] in

[1] J. M. Otero and A. Durán, Ann. Física y Química **37**, 459 (1941).
[2] P. Jiménez-Landi and J. Cabello, Ann. Física y Química **39**, 597 (1943).
[3] F. Schupfer, Boll. Ass. Ott. Ital. (January 1944).
[4] G. Wald and D. R. Griffin, J. Opt. Soc. Am. **37**, 321 (1947).
[5] V. Ronchi, Rev. Sapere **161**, 199 (1943).
[6] A. Durán, Ann. Física y Química **39**, 567 (1943).
[7] S. Hecht, J. Opt. Soc. Am. **32**, 42 (1942).

which quantum fluctuations of the threshold were displayed. Hecht and his collaborators established that 5, 6, or 7 absorbed quanta are needed to stimulate a minimal light perception. Other investigators[8] have found a still smaller number.

At the same time, if the fall in threshold depends upon the size of the object, it could imply some alterations in Riccó's law or, at least, in the limits within which it remains valid.

EXPERIMENTAL TECHNIQUE

a. Physical

The test object consisted in the aperture of an Ulbricht sphere, closing of which was carried out by means of diaphragms whose diameters had been carefully measured in a comparator. This device supplied us with a source of rigorously homogeneous brightness. The observer, in front of the sphere, and fitted with an appropriate chin-rest, was placed in such a way that the distance between his eyes and the aperture of the sphere was always 8 meters. The sphere was illuminated by a tungsten gas-filled incandescent lamp which had been previously aged. Its total flux was of the order of 242 lumens. The constancy of its working characteristics was assured by feeding it from a battery of accumulators of large capacity. Voltage and current were controlled, and the invariability of the working characteristics of the lamp was duly checked.

The spectral energy distribution corresponded to a color temperature of 2300°K. Reduction of brightness by each diaphragm was in two stages, the first employing grey, decadic filters which had been previously calibrated for the scotopic and photopic visual ranges in a König-Martens spectrophotometer. Fine adjustment was achieved through a rotary sector, care having been taken not to reduce the free sector to less than 5 percent of the whole aperture. The interposition of any supplementary filter was avoided during any single set of measurements, adjustments being made by means of the sector alone.

Figure 1 shows the plan of the assembled device.

b. Physiological

As the number of necessary measurements was very high, *viz.*, over 50,000 determinations, because of the statistical method employed, we were compelled to restrict ourselves to only four observers, three of them being the authors of this paper (a physician and two physicists) and the fourth an assistant in the Institute, who also was experienced in visual measurements. All of them were emmetropic and possessed visual acuities slightly above the standard. Their ages ranged between 24 and 40, and they all possessed normal accommodations, the eldest of them being able easily to accommodate 5 diopters.

Observation was binocular. Dark adaptation was never for less than 30 minutes in complete darkness. Fixation was accomplished with a small incandescent lamp, located inside a capsule. The latter was screened with a red gelatine

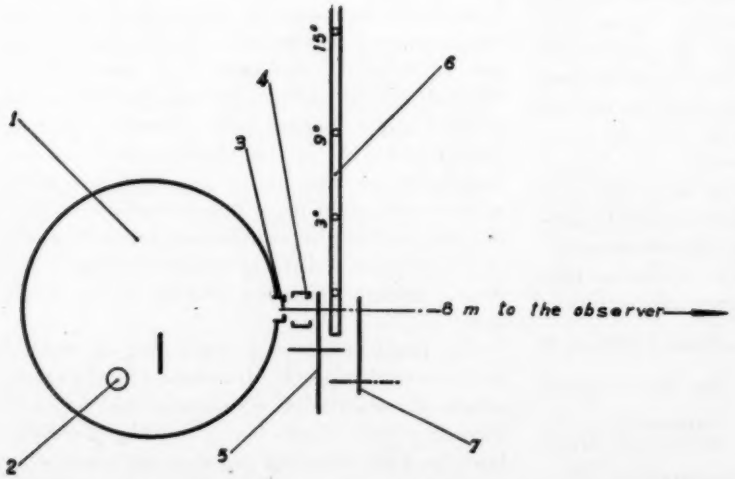


FIG. 1. Experimental device: (1) Ulbricht sphere; (2) lamp; (3) diaphragm; (4) filter holder; (5) rotating sector; (6) fixation lamp; (7) retractable screen.

[8] M. A. Bouman and H. A. van der Velden, J. Opt. Soc. Am. 37, 908 (1947).

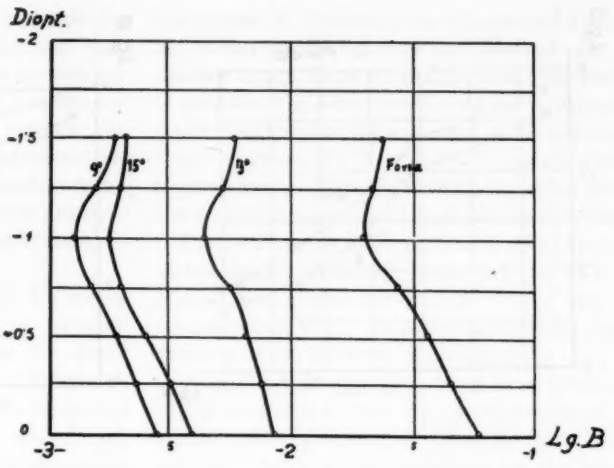


FIG. 2. Absolute threshold in millilamberts as a function of accommodation for the four retinal zones studied. Observer 2, diaphragm diameter 2.65 minutes.

filter, eliminating wave-lengths below 6400A. This lamp illuminated a small opal glass screen bearing the filter, and fitted with a diaphragm with which its apparent diameter could be altered at will.

For foveal as well as extra-foveal explorations, the fixation lamp could be placed at various positions along a board perpendicular to the observer-test line. These positions corresponded to foveal fixation and excentricities of 3°, 9°, and 15°.

While the foveal examination was carried out, the lamp remained completely contiguous to the aperture of the Ulbricht sphere in such a way that the angle, test-observer lamp, was smaller than 20 minutes, so securing pure foveal fixation.

Absolute thresholds were measured for test fields which subtended angles ranging from 30 sec. to 11 min. for the fovea and for retinal zones 3°, 9°, and 15° from the fovea.

In order to study the influence of night myopia upon the absolute threshold, the latter was measured for each size of field and retinal region, using Ostwalt spectacles whose power varied in steps of $\frac{1}{4}$ dioptrie from 0 to -1.75 or -2 dioptries. Their use made practically nil the possible perturbing influence of astigmatism appearing when exploring the extra-foveal zones up to 15°. By computing the curvature of the field of all the spectacles—which came out to be substantially equal—and by checking the results obtained with the naked eye against those attained with eyeglasses of 0 dioptrie, the possible perturbing influence of extra-axial aberrations was eliminated. The diaphragm fitted to the auxiliary small red lamp allowed for reducing its apparent size, thus compensating for the enlargement of the image due to being out of focus when seen through glasses of higher power. This lamp was also fitted with a rheostat to maintain its brightness at the appropriate value, and so not disturb

TABLE I. Absolute threshold in millilamberts, as a function of the size of test field, for the normal eye and for vision corrected for night myopia, together with the relative descent of the threshold and the number of dioptries needed to compensate the ametropia.

Field diam. (min.)	Absolute thresholds (millilamberts) Normal eye	Corrected for night myopia	Difference in threshold	Night myopia in dioptries
		Fovea		
10.30	0.0065	0.0038	58.4%	−0.85
08.50	0.0074	0.0054	73.0%	−1.00
05.10	0.0190	0.0071	37.4%	−1.05
02.65	0.0470	0.0220	46.8%	−1.00
01.80	0.2100	0.1100	52.4%	−1.00
00.50	0.6500	0.3600	55.4%	−0.75
		at 3° from the fovea		
10.30	0.00029	0.00015	51.7%	−1.05
08.50	0.00076	0.00041	53.9%	−1.00
05.10	0.00280	0.00140	50.0%	−1.00
02.65	0.00880	0.00440	50.0%	−1.00
01.80	0.03600	0.01900	52.8%	−1.00
00.50	0.10000	0.05200	52.0%	−1.10
		at 9° from the fovea		
10.30	0.00018	0.00011	61.1%	−0.80
08.50	0.00047	0.00027	57.4%	−1.00
05.10	0.00100	0.00061	61.0%	−0.85
02.65	0.00300	0.00140	46.7%	−1.05
01.80	0.01400	0.00690	49.3%	−0.85
00.50	0.04400	0.02600	59.1%	−0.85
		at 15° from the fovea		
10.30	0.00029	0.00013	44.8%	−0.80
08.50	0.00066	0.00036	54.5%	−1.05
05.10	0.00140	0.00079	56.4%	−1.20
02.65	0.00530	0.00240	45.3%	−1.10
01.80	0.03500	0.01800	51.4%	−1.00
00.50	0.08600	0.04400	51.2%	−1.00

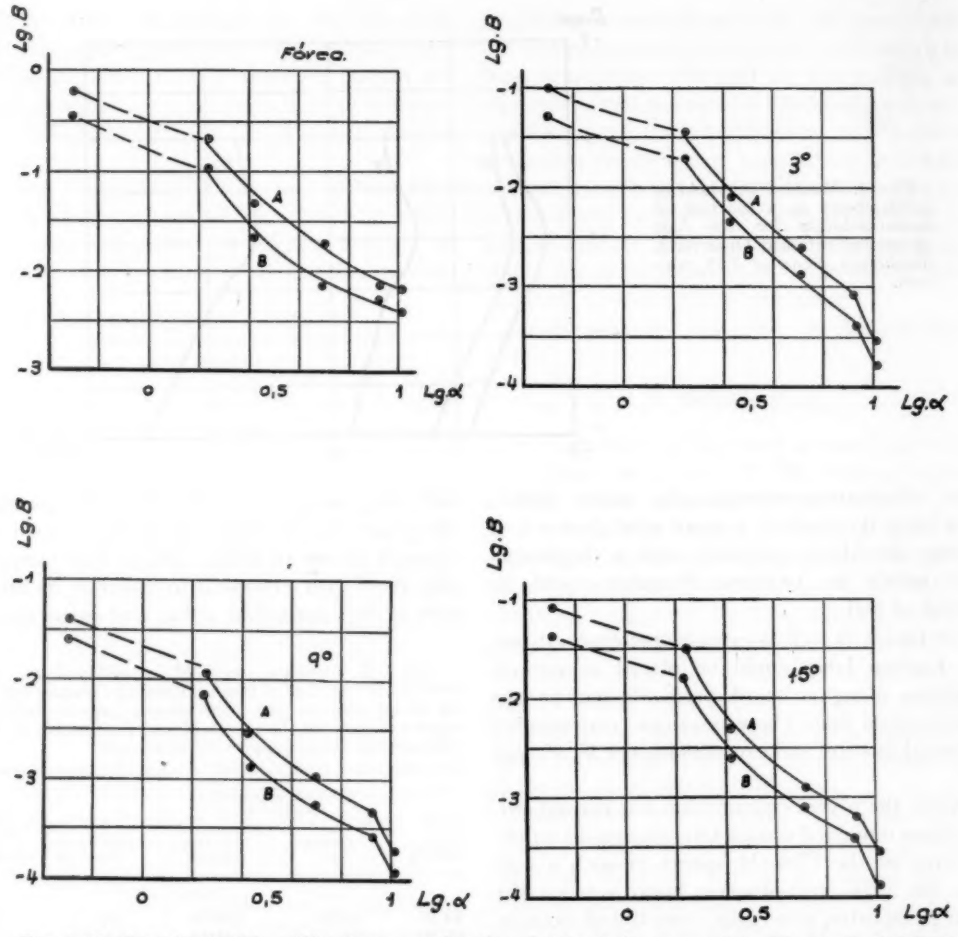


FIG. 3. Absolute threshold in millilamberts, as a function of the test size measured in minutes, for the fovea, and at 3°, 9°, and 15° of same. Curves *A* correspond to the normal eye and *B* to corrected night myopia. Ordinates: logarithm of the threshold brightness in millilamberts; abscissae: logarithm of the diameter of the test field in minutes.

the measurements by glaring at the lowest adaptation levels.

In making measurements we started at brightnesses high enough so that the test field was definitely visible. To make the measurements more objective an opaque screen was occasionally inserted over the field, and the observer was asked to report whether he saw light or not.

The threshold was taken to be the brightness at which the observer reported 60 percent correct answers to the question concerning the presence or absence of the stimulus. Whenever the percentage of correct answers went over 60 percent the brightness of the test field was slowly reduced until reaching it. This method turned out to be very accurate, yielding reproducible values with very small errors when operating at different times with the same flux. Observers were not made to give their answers within any fixed time.

RESULTS OBTAINED

Once the determinations were carried out, curves were drawn for each observer and each diaphragm representing the brightness threshold as a function of the lens power with which the observation was performed in each particular case. Figure 2 illustrates one of these representative curves; the remaining ones have analogous shapes but are displaced by various amounts on the abscissa.

From this set of curves the brightness thresholds were derived for normal as well as for com-

pensated nocturnal myopia, for each particular position and field size studied. One can also derive from them the power of lens needed for compensating nocturnal myopia. As a summary of the results achieved, Table I records the average values of these magnitudes obtained for each of the four observers: In the fourth column of the table the relative decrease of the absolute threshold when night myopia is compensated is also shown.

From this table it is inferred that for every field size and retinal position—and also for each particular observer—the correction of night myopia provokes a variable decrease in threshold, reducing it to values ranging between 60 and 40 percent of those obtained with the naked eye. This decrease is little influenced by the size of the test field. What small effects occur do not follow any accurate law. The greatest variations are met in the fovea, and it is necessary to bear in mind that in this instance they may be influenced by the red fixation lamp, which in this case is located close to the test patch. In the other three positions of the retina studied the variations due to the size of the test field are very small.

The largest average gain seems to be found in the zone 9° from the fovea. At the same time it can be seen that the greatest retinal sensitivity is found in this same zone, contrary to frequent reports that it is located somewhere between 15° and 30° from the fovea. It must not be forgotten that when binocular vision is involved, the excited retinal zones lie on opposite sides—nasal and temporal—of the fovea. But, as is well known, the regions of equal sensitivity of the retina are approximately zones of spherical surface, with their axis passing through the fovea, and so the excited places on both eyes had very close to the same sensitivity. This threshold minimum for the 9° zone was obtained for all the field sizes and all the observers.

The correction for night myopia at which the minimum threshold is attained ranges between $\frac{3}{4}$ and 1.25 dioptrie, but it is necessary to submit these figures to two further corrections. The first of 0.125 dioptrie is due to the fact that we operate at a distance of 8 meters, and therefore subject the eye to a slight accommodation which must be added to those figures arrived at with the correcting glasses. The second is brought about because, when using a red light as fixation point, we automatically introduce a displacement of focus of +0.5 dioptrie owing to the chromatic aberration of the eye, from the normal focusing by 5550A in daylight to the 6400 of the red fixation lamp. Thus we gather from this that the values directly determined must be increased by 0.625 dioptrie; and the actual values therefore range between 1.375 and 1.875, which, though slightly lower, agree with those obtained for the same observers using a visual acuity criterion.

Concerning the fulfilment of Ricc6's law in the visual angles explored (from 0.5′ to 11′), the results are illustrated in Fig. 3 for the four retinal zones which we have explored in normal vision and in corrected night myopia.

The first fact emerging from the examination of these curves is that the phenomenon of nocturnal myopia does not alter at all the law relating the absolute threshold to the area of the stimulus, at least in the range of areas studied.

The curves drawn agree completely with those given by Graham, Brown, and Mote[9] in the interval studied in both researches, both in the fovea and in extra-foveal zones. However, the beginning of the second branch of the extra-foveal function occurs at test diameters of about 10 minutes instead of the 20 minutes determined by the cited investigators. Also at test fields of 0.5 minute we obtain absolute thresholds smaller in every case than those expected if the curve and formulae given by Graham (which only reach up to 1.8 minutes) were to be extended to smaller test diameters. This seems to indicate the existence of a new branch of the curve (dotted line in the drawings) for very small field sizes, comparable with that appearing in the periphery for large fields. It can readily be seen that the three branches of the curve possess increasing slopes, the smaller slopes going with the smaller stimuli.

CONCLUSION

A 50 percent decrease in the absolute threshold has been found when nocturnal myopia is compensated, a drop which is perfectly explained if we bear in mind that all the luminous flux distributed in the diffuse image during normal observation is gathered into a much smaller surface when the night myopia is corrected, i.e., when the size of the image is reduced to that which diffraction and the geometrical optics of the eye determine. By this means the brightness of the retinal image is increased, surpassing the thresh-

[9] C. H. Graham, R. A. Brown, and F. A. Mote, J. Exper. Psych. 24, 555 (1939).

old and thus allowing for a reduction of the brightness of the stimulus to the new threshold, and this fact is independent of the time of observation. So this fall in threshold must happen also using flashes of light as in Hecht's and van der Velden's experiments.[7,8]

Therefore this fall in threshold may reduce the number of quanta needed to stimulate a minimal light perception to half the number set by those researchers who have experimented without considering this phenomenon.* As one consequence, if we accept van der Velden's results, only one quantum may be enough to evoke a light perception.

This same decrease in threshold shows that the correction of night myopia will enable us either to recognize luminous signals half the brightness of those normally perceived, or to do so at longer distances with a given brightness.

The law relating the absolute threshold to the size of the test field is not influenced by the night myopia phenomenon within the 0.5′–11′ in which Riccó's law is generally considered valid.[10–12] It is possible, however, that its zone of approximate validity extends up to larger angles once the nocturnal myopia has been compensated. We are endeavoring to explore this in our new experiments.

* One of us called at Professor Hecht's laboratory a few days after his death. He was able to see the apparatus with which the late Professor Hecht carried out his measurements of quantum fluctuations, but, unfortunately, Dr. Hsia, who showed him the laboratory, was unable to tell us the distance at which the fixation lamp was located, a factor affecting the state of accommodation of the eye. In any case it seems very improbable that these distances agree with those given for compensating night myopia.

[10] S. Löhle, Zeits. f. Physik **60**, 137 (1929).
[11] B. Schönwald, Das Licht **11**, 15 (1941).
[12] H. Richard, J. Opt. Soc. Am. **36**, 624 (1946).

JOURNAL OF THE OPTICAL SOCIETY OF AMERICA VOLUME 39, NUMBER 2 FEBRUARY, 1949

The Course of Foveal Light Adaptation Measured by the Threshold Intensity Increment*

HOWARD DEHAVEN BAKER**
Department of Psychology, Columbia University, New York, New York
(Received August 25, 1948)

The course of change of sensitivity during illumination of the fovea has been determined for various intensities of adapting light. Before illumination the fovea was in a dark-adapted condition. Sensitivity is defined as the reciprocal of the just perceptible intensity increment. It is found that, for all illuminations used, sensitivity, as defined above, is low immediately after the onset of the adapting light; it increases to a maximum several minutes after the onset of the adapting illumination and then declines to an intermediate final level. Some theoretical implications of this finding are considered.

LIGHT adaptation may be defined as the change in sensitivity of an initially dark-adapted eye which occurs as the eye is exposed to light. Studies of the course of light adaptation in the human eye have indicated that a previously dark-adapted eye is very sensitive during the first moments of exposure to light, but that as illumination continues, sensitivity decreases progressively (von Kries,[1] Lohmann,[2] Wright,[3] Geldard,[4] Wald and Clark[5]). In these studies sensitivity has been defined in two ways: (a) as the reciprocal of the absolute threshold determined at a given interval after the light-adapting light has been turned off (Lohmann,[2] Wald and Clark[5]), and (b) as the brightness, relative to the adapting light, of a light which matches the "apparent" brightness of the adapting light (Wright,[3] Geldard[4]). In general, researches on light

* Submitted in partial fulfillment of the requirements for the degree of Doctor of Philosophy in the Faculty of Pure Science, Columbia University.

Prepared under Project NR 142–404, Contract N6onr–271, Task Order IX, between Columbia University and the Office of Naval Research, U. S. Navy.

** Now at the Department of Psychology, Johns Hopkins University. The author wishes to express his sincere gratitude to Professor Clarence H. Graham, under whose direction the present research was accomplished, and to Professor W. N. Schoenfeld for many suggestions and a critical reading of the manuscript. He is also indebted to Mr. Alan Hoffman for assistance in deriving some of the necessary mathematical statements, and to Mr. Edwin W. Bechtold for apparatus suggestions.

[1] J. von Kries, "Ueber Ermüdung des Sehnerven," Graefe's Arch. f. Ophth. **23** (2), 1 (1877).
[2] W. Lohmann, "Ueber Helladaptation," Zeits. Sinnesphysiol. **41**, 290 (1906).
[3] W. D. Wright, "The measurement and analysis of colour adaptation phenomena," Proc. Roy. Soc. London **115B**, 49 (1934).
[4] F. Geldard, "The measure of retinal fatigue to achromatic stimulation," J. Gen. Psychol. **1**, 123, 578 (1928).
[5] G. Wald and A. B. Clark, "Visual adaptation and the chemistry of the rods," J. Gen. Physiol. **21**, 93 (1937).

adaptation have indicated that sensitivity, as defined in either of the above ways, decreases rapidly at first, than more slowly as an asymptotic level is approached. A nearly asymptotic sensitivity is said to represent complete adaptation to the incident illumination. The level of sensitivity finally reached is a function of the intensity of the adapting light.

The definitions of sensitivity referred to above are based on two different methods of measuring the course of light adaptation. The method which was employed, with various modifications in each study, by von Kries, Wright, and Geldard, involves exposing one portion of the retina of a dark-adapted eye to a constant or adapting illumination. An adjacent area is left dark-adapted. At various intervals the dark-adapted portion is then exposed to an illumination, which the subject matches in apparent brightness to the area which has been adapting to the constant illumination. It is found that under these circumstances the dark-adapted portion requires less physical brightness in order that it be subjectively matched with the light-adapted area. This is taken as evidence that the sensitivity has been reduced in the area exposed to light, and the remaining sensitivity is defined by the ratio: $\text{Sensitivity} = \frac{\text{brightness of dark-adapted field}}{\text{brightness of light-adapted field}}$. This ratio is usually stated in percent. As the adapting illumination continues, sensitivity decreases, thus revealing the course of the light adaptation process. However, no theoretical system, of the sort proposed by Hecht[6] or Jahn,[7] for example, has as yet been extended to include the measure of sensitivity utilized by this method.

The other method, employed (again with modifications) by Lohmann and by Wald and Clark, uses the reciprocal of the absolute threshold as its measure of visual sensitivity. This measure has the advantage at present that it is easily relatable to theories as they now exist (Hecht,[6] Jahn[7]), but it has the disadvantage that its determination during the course of light adaptation involves technical difficulties. The method requires measuring the absolute threshold at various times following the onset of the adapting light. The curve of the change in sensitivity is then constructed from the resulting data. In order to measure the absolute threshold it is necessary to have the subject in darkness. Consequently, the adapting light must be discontinued whenever a threshold is measured, and since the measurement of a threshold with standard adaptometers requires a period of several seconds, some dark adaptation must occur before each measurement. In view of the rapidity of the dark adaptation process at its beginning, the error in the measurement is not inconsiderable. It may be assumed that the error remains constant for threshold measures taken at identical periods in the dark, and hence that the course of light adaptation is truly represented by the curves except for a displacement on the time axis. This assumption must be accepted with reservations in view of the known difference in initial velocity of the dark-adaptation process following different periods of exposure to the same illumination and following exposure to different amounts of illumination (Wald and Clark[8]). Thus, light adaptation as measured by the absolute threshold method must be regarded as an approximation. Nevertheless, results by the method again indicate high sensitivity at the beginning of the process and a lower sensitivity later.

Although the shapes of light-adaptation curves given by both methods discussed are roughly similar, an examination of the results reveals some interesting quantitative differences. Lohmann and Wald and Clark find that light adaptation, examined by the threshold method and in terms of their defined sensitivity, requires from ten to eighty minutes for completion, whereas von Kries, Wright, and Geldard find, by the comparison method, that the process is complete in less than three. The total changes in sensitivity also vary, but since entirely different operations are involved in making the measurements, little emphasis should be placed on comparisons in the absence of theoretical relations.

The present experiment determines the course of cone light adaptation in terms of a criterion of visual sensitivity which may be related to a theoretical framework and which does not require darkness for its measurement. Ideally, light adaptation should be measured by instantaneous determinations of a theoretically significant variable at various intervals after illumination is begun, and on the basis of a method which does not alter the course of the process by its application. The present research endeavors to approximate such conditions. The significant variable studied is the value of the increment in intensity, ΔI, required to produce a just perceptible brightening of the visual field at various times after the onset of an adapting light. On the practical side, it is assumed that the procedure of presenting ΔI in discrete flashes, each of 0.02-second duration, at 1-second intervals does not seriously distort the course of the light-adaptation process. On the theoretical side, it is to be observed that Hecht's[8] equations for intensity discrimination imply the existence of a simple relationship between ΔI and the amount of the photosensitive material

[6] S. Hecht, "Rods, cones, and the chemical basis of vision," Physiol. Rev. 17, 239 (1937).

[7] T. L. Jahn, "Brightness discrimination and visual acuity as functions of intensity," J. Opt. Soc. Am. 36, 83 (1946).

[8] S. Hecht, "A theory of visual intensity discrimination," J. Gen. Physiol. 18, 767 (1935).

present in the photo-receptor during the action of light. The implications of our results for theory will be considered in a later section of the report.

All measures are taken in the rod-free area of the fovea.

APPARATUS AND PROCEDURE

The apparatus serves the function of presenting the subject with a uniform field of light, to which may be added, at spaced intervals, an additional illumination, ΔI, variable in intensity and limited to the fovea. The threshold ΔI may be determined by observing the value of ΔI which is just perceptible. The course of light adaptation is traced by determining the threshold ΔI at various times following the onset of the adapting illumination. White light is used throughout. The adapting field is made to subtend a visual field of twelve degrees, in order to provide a wide surround for the stimulus patch and minimize error due to eye movements. Central fixation is maintained by use of a black fixation point subtending five minutes of visual angle. The stimulus area of the added flash is

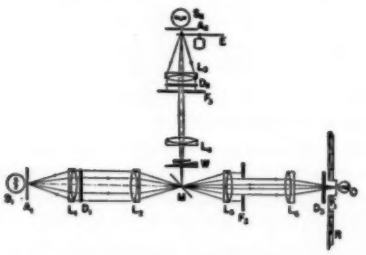


FIG. 1. Diagram of the apparatus. For explanation see text.

circular, and appears concentrically with the fixation point. The angle subtended by the area of the additional flash is one degree; hence the image of the ΔI area falls within the rod-free region of the fovea.

A schematic diagram of the apparatus appears in Fig. 1. A lens (L_1) collimates the light from a source approximating a point, obtained by placing a metal plate with a 1.85-mm aperture (A_1) before a tungsten filament light bulb (S_1). The parallel light emerging from L_1 is brought to a focus by a lens (L_2) upon a plate glass reflector (M). Some of the light passes directly through the reflector and is collimated by a succeeding lens (L_5). The parallel light is brought to a focus by an imaging lens (L_6) within the pupil of the subject's eye (O), which is positioned by an artificial pupil on an eyepiece (F_1). This provides an effective pupil area of 0.856-square millimeters. The eyepiece projects through a wall (R) into a darkroom in which the subject is seated. This optical system creates the so-called "Maxwellian view" which yields an evenly illuminated field of comparatively intense light upon the retina. The visual field is defined in the present apparatus by a field stop (F_2) at focal length from L_6, and includes all of the flux passing through the field stop. The field subtends twelve degrees at L_6. A plate glass over the field stop carries a small spot of India ink in its center; the spot subtends five minutes at L_6 and serves as a fixation point. This portion of the system provides the adapting field which, when once turned on, remains illuminated at a constant brightness throughout the course of the experiment.

The ΔI flashes are provided by an optical system entering the glass reflector at a right angle. A tungsten filament light source (S_2) provides the illumination for the 1.85-mm aperture (A_2), the light from which is collimated by a lens (L_3). A sector disk shutter (E), run by a synchronous motor, limits the illumination to flashes of 0.02-second duration, spaced at intervals of one second. The parallel light leaving L_3 is limited by a field stop (F_3), at focal length from lens (L_4), which regulates the size of the ΔI area, controlling the subtense to one degree of visual angle at the eye. The intensity of the stimulus beam is controlled by a Wratten neutral tint gelatin wedge with a balancer (W). The light from the stimulus beam is imaged through the wedge upon the plate glass reflector (M), and some of the light is reflected into the identical course of the beam of the adapting field. Thus the ΔI beam strikes the retina of the subject under the same conditions as the adapting field and is concentric with the fixation point; it appears as a spot of added brightness one degree in diameter upon the center of the 12-degree field of the adapting beam.

The intensity of the two beams can be controlled in stages by means of Wratten neutral tint gelatin filters at points indicated by D_1, D_2, and D_3 on the diagram, and the stimulus spot can be adjusted to a continuous range of brightness by means of the wedge (W). The filament lamps are operated on direct current. All intensity changes made during the experiment are made by altering the total optical density of the wedge and filters.

The brightness of the fields of light is calibrated by binocular match with an evenly illuminated surface viewed through an artificial pupil of the same size as that provided by the apparatus. With no filters in the beam of the adapting field, its apparent brightness is equal to that of a surface whose brightness (as measured by a Macbeth illuminometer) provides a retinal illumination of 50,000 photons. The measure of retinal illumination was calculated by the formula: photons $= (10/\pi) \times$ pupil area in square millimeters $\times$ brightness in millilamberts. The adapting field thus presents a surface of 5800-millilamberts brightness viewed

through an effective pupil area of 0.856 mm². The stimulus spot gives an apparent brightness, determinable within the limits of accuracy of a binocular comparison, which matches a surface yielding a retinal illumination of 20,000 photons, or a brightness of 2300 millilamberts viewed through a pupil area of 0.856 mm².

The determination of threshold ΔI employs the method of limits, modified to include the ascending series only. (See the description of procedure which follows immediately below.) The use of a descending series would require supraliminal flashes, with an attendant danger of distorting the course of light adaptation.

In the experiment the subject is seated in the darkroom, and is kept in darkness for an initial period of ten minutes. This period was found, in preliminary determinations, to be sufficient to give complete cone dark adaptation. Upon the completion of dark adaptation, the subject is given a warning signal to place his eye at the eyepiece which projects through the wall of the darkroom. When the subject is ready to observe, the adapting field is turned on and the sector disk shutter is

TABLE I. Average data of subject D.Y., based upon a number of determinations (N) of the log threshold intensity increment (ΔI_t) at various times (t) after onset of the adapting light of four intensities.

5000 photons (580 millilamberts; pupil area 0.856 mm²)			500 photons (58 millilamberts; pupil area 0.856 mm²)		
N	time (seconds)	logΔI_t (photons)	N	time (seconds)	logΔI_t (photons)
8	5	3.42	8	5.25	2.28
11	23	3.15	13	20.2	2.04
10	37	3.06	9	36.8	1.93
7	53	2.99	9	52.7	1.88
17	74	2.96	14	85.0	1.84
14	105	2.96	13	102	1.88
16	133	2.97	14	135	1.86
15	161	2.90	17	165	1.86
16	195	2.90	14	196	1.85
9	219	2.93	13	225	1.85
24	265	2.99	25	267	1.87
23	321	3.00	24	325	1.87
24	384	2.98	22	385	1.87
22	444	3.04	24	448	1.88
22	505	3.06	22	503	1.89
23	562	3.04	24	566	1.92
24	627	3.01	23	626	1.90
25	687	3.02	24	686	1.89
23	747	3.05	24	744	1.88
24	805	3.08	24	806	1.90
21	867	3.06	23	868	1.90
18	918	3.07	24	922	1.89
50 photons (5.8 millilamberts; pupil area 0.856 mm²)			5 photons (0.58 millilamberts; pupil area 0.856 mm²)		
N	time (seconds)	logΔI_t (photons)	N	time (seconds)	logΔI_t (photons)
9	3.11	1.82	8	3.38	0.61
11	19.9	1.48	13	19.9	0.51
14	44.1	1.37	8	37.5	0.48
15	72	1.38	7	52.9	0.45
13	101	1.34	13	72.9	0.43
18	131	1.31	17	103	0.44
14	161	1.30	16	133	0.43
15	193	1.28	16	163	0.43
15	222	1.27	15	194	0.43
25	261	1.28	15	226	0.45
23	324	1.32	23	266	0.47
24	385	1.32	24	320	0.44
24	445	1.34	25	384	0.46
24	504	1.32	24	447	0.46
24	563	1.33	24	507	0.48
24	623	1.37	24	568	0.48
23	682	1.37	25	628	0.48
24	746	1.35	23	689	0.50
24	805	1.38	23	747	0.49
22	862	1.37	23	808	0.51
24	920	1.39	24	866	0.51
			23	920	0.49

TABLE II. Average data of subject A.B. based upon a number of determinations (N) of the log threshold intensity increment (ΔI_t) at various times (t) after onset of the adapting light of four intensities.

5000 photons (580 millilamberts; pupil area 0.856 mm²)			500 photons (58 millilamberts; pupil area 0.856 mm²)		
N	time (seconds)	logΔI_t (photons)	N	time (seconds)	logΔI_t (photons)
7	4.7	3.42	8	4.50	2.21
8	19.6	3.26	11	20.0	2.06
9	37.7	3.16	9	35.4	1.99
7	54.1	3.09	10	51.9	1.95
14	75.0	2.98	18	75.6	1.90
13	104	2.93	16	106	1.87
15	134	2.89	15	136	1.88
13	164	2.90	16	164	1.87
12	197	2.90	16	198	1.87
14	226	2.95	16	225	1.88
24	271	2.98	26	269	1.88
25	329	3.01	27	326	1.88
22	388	3.04	25	388	1.89
25	451	3.09	28	449	1.89
24	507	3.07	25	508	1.90
24	567	3.06	27	567	1.91
22	627	3.11	27	626	1.91
22	687	3.10	27	687	1.92
23	748	3.09	24	745	1.93
22	807	3.07	26	806	1.93
24	868	3.12	27	875	1.94
22	928	3.11	24	919	1.95
50 photons (5.8 millilamberts; pupil area 0.856 mm²)			5 photons (0.58 millilamberts; pupil area 0.856 mm²)		
N	time (seconds)	logΔI_t (photons)	N	time (seconds)	logΔI_t (photons)
8	4.62	1.60	8	3.62	0.57
13	19.8	1.40	13	20.5	0.45
9	36.9	1.37	9	36.5	0.44
8	51.0	1.36	9	51.6	0.42
16	76.0	1.35	18	73.0	0.39
13	104	1.31	16	102	0.37
18	132	1.30	16	134	0.38
14	161	1.32	16	163	0.38
17	192	1.33	16	192	0.37
14	221	1.33	15	222	0.39
24	261	1.32	25	266	0.39
24	322	1.33	25	323	0.40
22	383	1.31	25	385	0.41
24	445	1.36	24	443	0.41
24	505	1.34	24	503	0.42
24	564	1.34	24	564	0.42
21	621	1.35	24	623	0.43
24	684	1.36	24	685	0.41
24	743	1.38	24	745	0.43
24	803	1.39	26	805	0.42
24	863	1.40	24	865	0.44
23	918	1.39	23	921	0.44

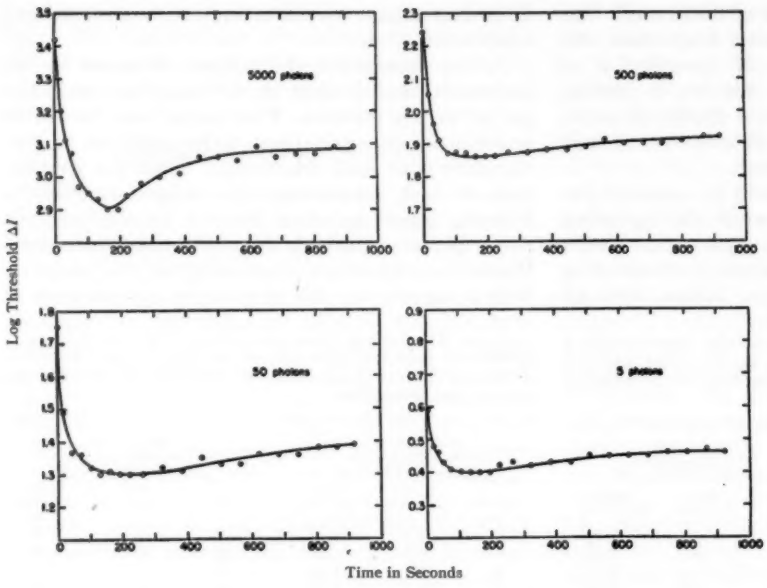


FIG. 2. Logarithm of the threshold intensity difference (in photons) as a function of time of exposure to four adapting intensities. The effective pupil area is 0.856 mm², from which photons = 8.56 × millilamberts. The data shown are averages of the data of two subjects.

started, so that at 1-second intervals a 1-degree centrally located flash of light, ΔI, of 0.02-second duration is added to the adapting field. By means of the wedge, the brightness of ΔI is increased at each flash until a momentary increase in illumination can be perceived by the subject. The subject signals this event by means of a telegraph key in the darkroom, which lights a signal lamp near the experimenter. The intensity of threshold ΔI and the time of presentation of the flash are recorded. The brightness of the flashes is then reduced again below threshold and the measurement repeated. In this way it is possible to record approximately three determinations of ΔI during each minute of exposure. The adaptation process is measured for a period of fifteen minutes. If a second experimental session is contemplated for the same day, the subject is again placed in the dark.

Two subjects were used, D.Y. and A.B., both women with normal vision. With each subject four adapting intensities were employed: illuminations of 5000, 500, 50, and 5 photons, respectively. Eight sessions were made at each intensity with each subject; each datum reported is, therefore, an average of several readings. The number of readings contributing to each datum varies from seven to twenty-eight for each subject, with greater numbers occurring in the data for longer adaptations. The difference in number of intensity readings is a consequence of the procedure necessitated by the rapidity of change in the early part of the process as contrasted with the slower effects in the latter part. Readings in the early part of adaptation had to be made at similar times in the different sessions if alterations in the apparent course of the process were to be avoided. In the experiments, the readings taken during the first minute of the eight sessions at each intensity were averaged over quarter-minute intervals, a procedure yielding four points for the first minute. The resulting data are based on single readings in each session, and the average for each fraction of a minute represents eight determinations for each subject. During the second to fourth minutes, inclusively, averages were taken for the first and last halves of each minute, and so the number of determinations entering each datum increases. For the rest of each session, data were averaged by minutes. It is to be noted that a procedure involving the averaging of points between certain limits is essentially a curve smoothing technique which assumes that the course of the process may be considered linear between the limits chosen, and which actually results in a slight reduction in the rates of change revealed by the determinations. No attempt is made to assess such an error, since the effect is probably small with respect to the variability encountered.

RESULTS

The data for each adapting intensity are presented in Table I for subject D.Y., and in Table II for subject A.B. In the tables, the *time* column refers to the number of seconds elapsed since onset of the adapting light at the moment the determination is made, the ΔI_t column refers to the threshold ΔI in log photons, and the N column refers to the number of individual determinations upon which the average datum is based.

The average data for both subjects are shown graphically in Fig. 2. Threshold values of ΔI are plotted on the vertical axis, and the times of the measurements are plotted on the horizontal axis. ΔI_t is large shortly after the adapting light first strikes the dark-adapted eye, and a high intensity of ΔI is required. During the first few minutes ΔI_t drops rapidly to a minimum after about three minutes and then increases again. ΔI_t subsequently "levels off" above the three-minute minimum, and the "leveling off" occurs after about ten minutes of adaptation at the highest adapting intensity. It may be observed in the curves for lower adapting intensities that the approach to the final level comes later as adapting intensity is decreased.

As the adapting intensity decreases, the range of variation in $\log\Delta I_t$ decreases also, the initial drop in threshold being the outstanding feature of the process represented by each curve. The range of change in $\log\Delta I_t$ is greatest with the highest adapting intensity, least with the lowest, and the two middle intensities exhibit an effect lying between the extremes. It is possible that the 50-photon curve exhibits a greater range in $\log\Delta I_t$ than the 500-photon curve, but the difference is small and, in general, the trend as outlined seems to be consistent.

The similarity of results given by the individual subjects may be judged in Fig. 3, which represents individual data for the 500-photon intensity. This intensity is chosen for the comparison because it reveals the greatest difference obtained with respect to curve shape; differences affecting curve shape are most accentuated at long adaptation times.

All curves show individual differences with regard to absolute values of ΔI_t, but curve shapes are little affected.

DISCUSSION

Visual sensitivity as defined by both the absolute threshold method and the comparison method has been found to decrease during light adaptation from a maximum immediately after the onset of the adapting illumination to a final asymptotic level. A third measure of sensitivity has also been used, i.e., the reciprocal of the intensity difference threshold, $1/\Delta I_t$ (cf., e.g., Holway and Hurvich[9]). The application of the single term "sensitivity" to all three of these measures is perhaps unwise, since we have no basis for expecting that they should behave in the same way as functions of comparable variables. At any rate, it is a fact that $1/\Delta I_t$ does not behave during light adaptation in a manner comparable to sensitivity obtained by the absolute threshold and comparison methods. Figure 2 reveals that ΔI_t manifests an initial drop after the onset of the adapting light to a minimum at about three minutes; thereafter the curve rises again. Sensitivity as defined by $1/\Delta I_t$ must therefore be low at first, then rise to a maximum, and drop off again to a final level

Although the results of this experiment are not in line with light-adaptation data obtained in the human eye by the threshold and comparison light methods, they are in accord with results obtained by Riggs and Graham[10] on the eye of *Limulus*. In their measurement of light adaptation in a single photo-receptor unit of *Limulus*, Riggs and Graham used a method similar to that followed in the present experiment. They found that sensitivity, as defined by the reciprocal of ΔI required to produce a constant added number of impulses, was low at first, then rose to a maximum and finally returned to a low final level. Plotted with the logarithm of time on the horizontal axis and the logarithm of sensitivity ($1/\Delta I_t$) on the vertical axis, their curve shows a rise and fall, with a maximum at about eighty seconds after onset of the adapting light.

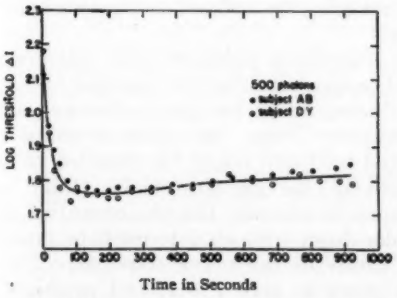


FIG. 3. Logarithm of the threshold intensity difference (photons) as a function of time of exposure to a retinal illumination of 500 photons (58 millilamberts; pupil area =0.856 mm²). The dates of the two subjects are plotted separately to show differences.

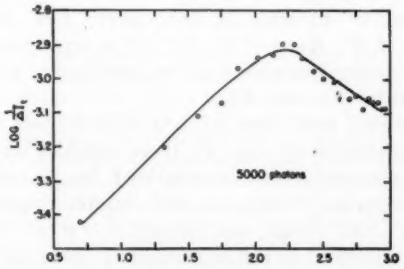


FIG. 4. The logarithm of sensitivity, as defined by the reciprocal of the threshold intensity difference, plotted as a function of the logarithm of time. The curve is for a retinal illumination of 5000 photons (580 millilamberts; pupil area =0.856 mm²).

[9] A. H. Holway and L. M. Hurvich, "Visual differential sensitivity and retinal area," Am. J. Psychol. 51, 687 (1938).

[10] L. A. Riggs and C. H. Graham, "Some aspects of light adaptation in a single photo-receptor unit," J. Cell. and Comp. Physiol. 16, 15 (1940).

Reference to Fig. 4 confirms the general formal similarity between the results of the present experiment and those of Riggs and Graham.

It is proposed in discussing the data of this experiment to consider their relation to some theoretical predictions. It will be worth while initially to consider the data in terms of Hecht's model of photo-reception, which is simpler than some later modifications. As will be seen, Hecht's simple account does not describe the data, and attention will be paid to more complex expectations based on other considerations.

If we follow Hecht[6] and consider the visual process as initiated by the formation of photo-products during irradiation of a photosensitive substance in the receptor, we can describe the rate of formation of photo-products as

$$dx/dt = k_1 I(a-x)^m - k_2 x^n, \tag{1}$$

in which x is the concentration of the photo-products; a is the initial concentration of the photosensitive substance at any particular time; k_1 and k_2 are the velocity constants of the forward and back reactions, respectively; and m and n refer to the order of the reactions. If we accept a value of 2 for both m and n, as dictated for the human fovea by the most pertinent data (Hecht[8]), and integrate Eq. (1), we find that

$$x = \frac{K_3 - K_1}{K_2} - \frac{2K_1}{K_2[(e^{K_1 t}/c') - 1]}. \tag{2}$$

In this equation the constants are defined as follows: $K_1 = 2ak_1k_2I$; $K_2 = 2(k_1I - k_2)$; $K_3 = 2k_1Ia$; and $c' = (K_3 - K_1/K_3 + K_1)$. The derivative of Eq. (2), which is equivalent to Eq. (1), can never equal zero for any finite value of t; hence there are no maxima or minima in the curve. For positive values of K_1, K_2, and K_3, Eq. (2) is represented by an increasing function with an asymptotic value of x equal to $(K_3 - K_1/K_2)$.

Consider, now, how ΔI acts with respect to the process implied by Eq. (2). If we consider intensity discrimination to be determined by a constant difference, $\Delta x_e (= \Delta x_{I+\Delta I} - \Delta x_I)$, in photo-products due to I and $I+\Delta I_t$ and formed in a short time, τ, we may state that the increment in photo-products due to the added intensity, ΔI_t, is

$$\Delta x_e = k_1 \tau \Delta I_t (a-x)^2, \tag{3}$$

where τ is less than a critical duration, τ_e (of the order of 0.05 second), and where the back reaction in the process due to $I+\Delta I$ is considered approximately equal to the back reaction in the process due to I. This is a generalized interpretation of Hecht's[8] analysis which applies to the stationary state condition; see also Graham and Kemp.[11] The analysis may mean that the subject notes a momentary difference, Δx_e, between the photo-products due to $I+\Delta I$, when ΔI is presented, and those due to I in the moment before or after ΔI is applied. If τ is constant, Eq. (3) may be written

$$\Delta I_t = [K/(a-x)^2], \tag{4}$$

where $K = (\Delta x_e / k_1 \tau)$. From Eq. (4) it appears that ΔI_t is reciprocally related to the concentration of the photosensitive substance at any time, and should therefore reflect the course of change in concentration. Thus, *sensitivity*, as defined by $1/\Delta I_t$, is theoretically related to the concentration of the photosensitive material.

Since Eq. (2) implies that x increases continuously with t to a limit at the stationary state, the behavior of $(a-x)^2$ is determined: it must decrease continuously during light adaptation. Therefore Hecht's ideas as extended in Eq. (4), which relates ΔI_t to $(a-x)^2$, require that ΔI_t increase continuously during light adaptation to the limit of the time required for the development of a stationary state. Clearly, the present data are not in accord with this expectation. ΔI_t does not rise continuously as time increases. Rather, it passes through a minimum at about three minutes.

Since the prediction derived from an extension of Hecht's theory to the present observations does not fit the data, other possibilities must be considered. In general, two points of view may be taken. The first is to assume that Hecht's simple theory fails to fit the data because the limiting factor in the determination of the response is not the photo-chemical process in the receptor but is, rather, the ability of the individual sense cell to respond. Thus, with an initial high rate of response to I, the sense cell may be approaching the response limit, and large values of ΔI are required to produce additional effects, until the rate of formation of photo-products slows to a point where that rate becomes a limiting factor. Beyond this point the curve reveals the course of the photo-chemical process.

The alternative point of view adheres to the original assumption that the limiting factor in the subject's response is the photo-chemical change in the receptor. Thus, the photo-chemical system suggested by Jahn[7] might be extended to describe qualitatively the present results. Jahn assumes that, upon irradiation, the photosensitive material S breaks down into an intermediate product, E, which catalyzes the nerve response; E, in turn, breaks down to give $P+A$, end products which recombine thermally to complete the reversible

[11] C. H. Graham and E. H. Kemp, "Brightness discrimination as a function of the duration of the increment in intensity," J. Gen. Physiol. 21, 635 (1938).

photo-chemical reaction. The total cycle of the process may be represented in the following paradigm: Light$\rightarrow S \rightarrow E \rightarrow (P+A) \rightarrow S$.

If we assume that the velocity constant of the reaction $S \rightarrow E$ is small with respect to the velocity constant of the reaction $E \rightarrow (P+A)$, it is possible to construct curves describing the concentration of E which show a maximum some time after the onset of the illumination on the previously dark-adapted receptor (Fry and Alpern[12]). A quantitative description of the course of human light adaptation from this point of view must await evaluation of constants. That the process of light adaptation involves considerable complexity in the matter of intermediate processes has been shown by the work on the nature of the visual cycle (Wald,[13] Wald and Clark[5]), and recent researches on dark adaptation and light adaptation in the single sense cell (Hartline and McDonald[14]) demonstrate that similar considerations apply to this case as well.

SUMMARY

1. The course of foveal light adaptation was measured by observing the way in which the threshold increment in light intensity, ΔI_t, varies, in two previously dark-adapted subjects, at various times during exposure to an adapting light, I. Four adapting intensities, ranging from 5 to 5000 photons, were used. The measurement of light adaptation in terms of ΔI_t has the advantage that the course of adaptation is little influenced by the addition of the small intensity increment, ΔI. In addition, the findings obtained by this method are relatable to known visual theory.

2. An apparatus for presenting brief flashes of ΔI upon the bright background, I, is described.

3. Sensitivity, as defined by the reciprocal of the threshold intensity difference, $1/\Delta I_t$, is, at all adapting intensities used, low immediately after the onset of the adapting light; it rises to a maximum after about three minutes and, subsequently, falls to a final intermediate level. ΔI_t is small for low values of I and increases as I increases.

4. The results are not comparable to results obtained by some earlier experimenters, who used different measures of sensitivity.

5. The experimental data are considered in relation to theories of photo-reception and neural processes in vision. The fact that sensitivity reaches an early maximum does not accord with simple accounts of the visual cycle. Consideration must be given to factors which complicate a simple interpretation of light adaptation.

[12] G. A. Fry and M. Alpern, "Theoretical implications of the response of a photo-receptor to a flash of light," Am. J. Optom. **21**, 1 (1946).
[13] G. Wald, "Visual systems and the vitamins A," Biol. Symp. **7**, 43 (1942).
[14] H. K. Hartline and P. R. McDonald, "Light and dark adaptation of single photo-receptive elements in the eye of Limulus," J. Cell. and Comp. Physiol. **30**, 225 (1947).

JOURNAL OF THE OPTICAL SOCIETY OF AMERICA VOLUME 39, NUMBER 2 FEBRUARY, 1949

Silicon Monoxide Protected Front-Surface Mirrors*

GEORG HASS AND NOEL W. SCOTT
Engineer Research and Development Laboratories, Fort Belvoir, Virginia
(Received October 11, 1948)

An evaporation method is described for producing front-surface mirrors with good abrasion, corrosion, and reflection qualities. Aluminum is used as the reflecting material because evaporated aluminum films have high reflectivity in all useful spectral ranges and show, in the electron microscope, a much finer grain and smoother surface than similar silver coatings. Since direct evaporation of SiO_2 is difficult and yields loose layers, a lower oxide of silicon, silicon monoxide, was investigated and found to produce good quality protective films. Silicon monoxide evaporates at a much lower temperature than silicon or silicon dioxide and deposits on the mirror surface in uniform and adherent layers which, when exposed to the atmosphere, partially oxidize to SiO_2. The visual reflectivity of aluminum mirrors protected with an optimum thickness (about 1500A) of SiO is about 89 percent, or only one percent lower than that of the unprotected aluminum surface. In the infra-red from 2μ to 8μ these layers do not show any appreciable absorption; however, the absorption band characteristic of silicon oxides appears between 8μ and 10μ. In the ultraviolet the SiO layer shows absorption which varies with the evaporation conditions. This absorption can be decreased by an oxidizing treatment. The SiO-protected mirror is unaffected by prolonged heating in air at 400C, and is highly resistant to sea water, sodium hydroxide, and most acids. The SiO-protected mirror also shows good abrasion resistance. To produce mirrors on metallic bases it is advisable to use an SiO film as an intermediate layer because it improves adherence and prevents diffusion from base metal to reflecting layer. With the SiO evaporation method it is also possible to make protected front-surface mirrors on plastic materials. Evaporated SiO layers are also successfully used as replica and support films in the electron microscope.

FRONT-SURFACE mirrors are almost universally produced by the high vacuum deposition method which has been adequately described in literature.[1] Since silver and aluminum have the

* Presented at the meeting of the Optical Society of America, New York, New York, October, 1946.

[1] John Strong, *Procedures in Experimental Physics* (Prentice-Hall, Inc., New York, 1944), pp. 151–187.

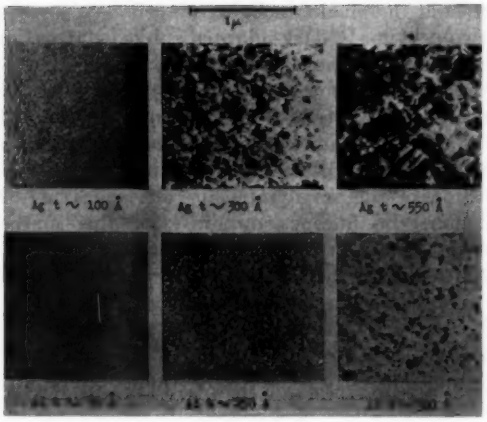


FIG. 1. Electron micrographs of Ag and Al films condensed to various thicknesses at room temperature.

highest reflectivity in the visible spectrum, they are the most frequently used mirror metals. Although silver has the higher visual reflectivity of the two, there are several factors that favor aluminum as the more desirable metal for producing efficient durable protected front-surface mirrors by the high vacuum deposition process.

I. THE MICROSTRUCTURE OF SILVER AND ALUMINUM MIRROR COATINGS

A study of the structure of thin evaporated films has been made by means of the electron microscope. The electron micrographs reproduced in Fig. 1 present a comparison of silver and aluminum films condensed in different thicknesses under good vacuum conditions (less than 10^{-4} mm Hg) on substrates at room temperature. For both metals, the graininess, and therefore the roughness, of the film surface increase with film thickness. For equal film thicknesses, the silver surfaces are much rougher than those of aluminum. Figure 2 shows three thicknesses of silver films condensed on bases at a temperature of 300C. A comparison of the silver films of Fig. 1 with those of Fig. 2 indicates that elevating the temperature of the substrate causes much rougher and more granular silver films. A similar result is shown for aluminum films in Fig. 3. In practice, the heating of the mirror form must be avoided when depositing silver as a reflector material, and the temperature of the mirror form when depositing aluminum should not be greater than 150C to avoid harmful roughness.

Since the extinction coefficient of silver is much smaller than that of aluminum,[2] thicker films of silver than of aluminum are required for opacity. Aluminum requires a film thickness of about 600A to be opaque, while silver films must be thicker than 1000A. It follows, therefore, that aluminum reflecting surfaces may be produced having smoothness approximating that of the aluminum film of 500A thickness shown in Fig. 1, while silver reflecting surfaces will be rougher than the silver film of 550A thickness shown in the same figure. The greater smoothness of evaporated aluminum surfaces, the resulting minimization of diffuse reflection, the well-known facts that evaporated aluminum films adhere to glass with greater tenacity than silver films and do not tarnish as easily, make aluminum a more desirable reflecting material for front-surface mirrors than silver, in spite of the higher reflectivity of the latter.

II. SILICON MONOXIDE AS MATERIAL FOR PROTECTIVE COATINGS

Although a thin protective film of aluminum oxide is formed on aluminum when it is exposed to air (thickness about 50A after one month),[3] evaporated aluminum coatings are mechanically and chemically too delicate for many mirror applications. Therefore, it is necessary to coat front-surface

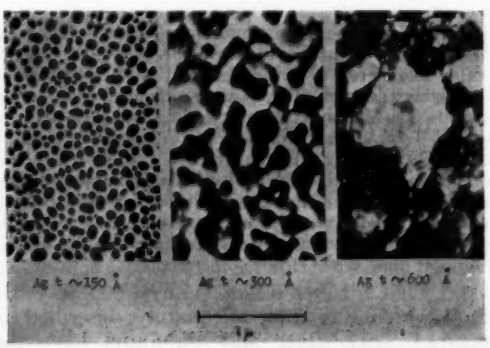


FIG. 2. Electron micrographs of Ag films condensed to various thicknesses at a temperature of 300°C.

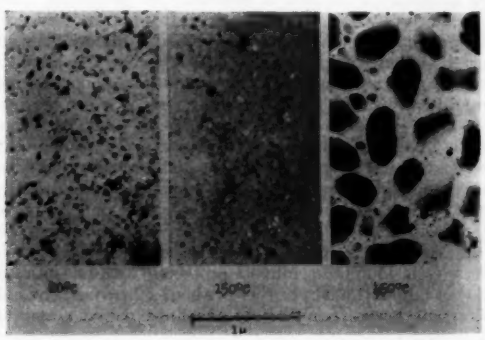


FIG. 3. Electron micrographs of Al films ~500A thick condensed at various temperatures.

[2] G. Hass, Optik 1, 2 (1946).
[3] G. Hass, Zeits. f. anorg. allgem. Chemie 254, 96 (1947).

aluminum mirrors with a transparent, durable material.

Because of its excellent physical properties, the use of silicon dioxide was considered for protective films. Silicon dioxide can be evaporated directly onto a reflector surface, but the evaporation is difficult and the resulting films have doubtful protective qualities because of loose structure and inadequate adhesion to the mirror material.

A lower oxide of silicon, silicon monoxide, has been found to evaporate easily and to produce hard adherent protective films for front-surface mirrors. Results of investigations made on this compound were first published in 1907 by Potter,[4] who patented several methods for producing silicon monoxide. More recent investigations have been made by Biltz and P. Ehrlich,[5] and Zintl.[6] The processes which produce silicon monoxide at relatively low temperature are as follows:

(1) Heating SiO_2 in an atmosphere of hydrogen:

$$SiO_2+H_2\rightarrow SiO+H_2O.$$

(2) Heating the following in a vacuum of the order of 1×10^{-4} mm Hg:

(a) $SiO_2+Si\rightarrow 2SiO.$

(b) SiO_2 in graphite crucible:

$$SiO_2+C\rightarrow SiO+CO.$$

(c) Si in metal oxide crucible:

$$Si+MeO\rightarrow SiO+Me.$$

(d) Bonded silicates with Si:

$$Al_2O_3\cdot 2SiO_2+2Si\rightarrow Al_2O_3+4SiO.$$

Several experimenters have made evaporations by placing silicon dioxide in graphite crucibles or silicon in metal oxide crucibles and have assumed that the condensed films consisted of silicon dioxide, or silicon, respectively. This assumption is not correct, because reactions in accordance with Eqs. (2b) and (2c) take place in the crucibles resulting in evaporated products which are mostly silicon monoxide. Electron diffraction studies have verified the above.

For producing quantities of silicon monoxide in the laboratory, reaction (2a) is suitable. Quantities of very finely divided silicon and silicon dioxide are intimately mixed approximately in proportion to their molecular weights. This mixture is placed in the closed end of a silica combustion tube. The tube is then evacuated to about 1×10^{-4} mm Hg, and inserted horizontally in a muffle furnace. Beginning at approximately 1100C, silicon monoxide evolves from the heated mixture and condenses on the cooler parts of the tube surfaces. The fused silica tube can be used but once, for it recrystallizes after being heated to 1200C.

The mixture of silicon dioxide and silicon can be used directly in the evaporation chamber, by heating in the evaporation boat, to produce protective films of silicon monoxide. However, more consistent results can be obtained by evaporation of silicon monoxide prepared in the above described manner.

The structure of thin evaporated films of silicon monoxide was studied by electron diffraction. Figure 4, an electron diffraction pattern of an evaporated silicon monoxide film of approximately 600A thickness, shows only one diffuse ring with d, the periodic distance between atoms, as calculated by means of Bragg's equation ($2d\sin\theta=n\lambda$), to be approximately 3.65A, which lies between the first

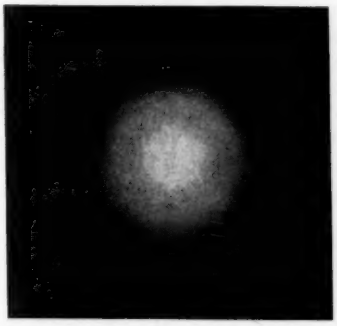


FIG. 4. Electron diffraction pattern of SiO film, approximately 600A thick.

pattern of silicon ($d=3.15$A) and that of silicon dioxide (d approximately 4.15A). Therefore, this material, which has neither the structure of silicon dioxide nor that of silicon, is a homogeneous amorphous material, the chemical formula of which is SiO. Its index of refraction varies with evaporation conditions and is 1.5 to 1.6 for good protective films. Evaporated silicon monoxide films partially oxidize to silicon dioxide when exposed to air. This oxidation stops after a very thin layer of silicon dioxide (about 100A thick) is formed on the silicon monoxide. A film of silicon monoxide 1500A thick is not completely oxidized after 100 hours in air at 400C. A more thorough oxidation takes place in steam at higher temperatures.

The relatively low temperature at which silicon monoxide can be evaporated, its amorphous structure, and the ability of thin films of the material to withstand high temperature also make it especially suitable for producing replica and support films for electron microscopy.

[4] Henry Noel Potter, Trans. Am. Electrochem. Soc. 12, 191 (1907), and U. S. Patent Nos. 875,286; 875,675; 875,676; and 866,636.

[5] W. Biltz and P. Ehrlich, Naturwiss. 26, 188 (1938).

[6] E. Zintl, Zeits. f. anorg. allgem. Chemie 245, 1 (1940).

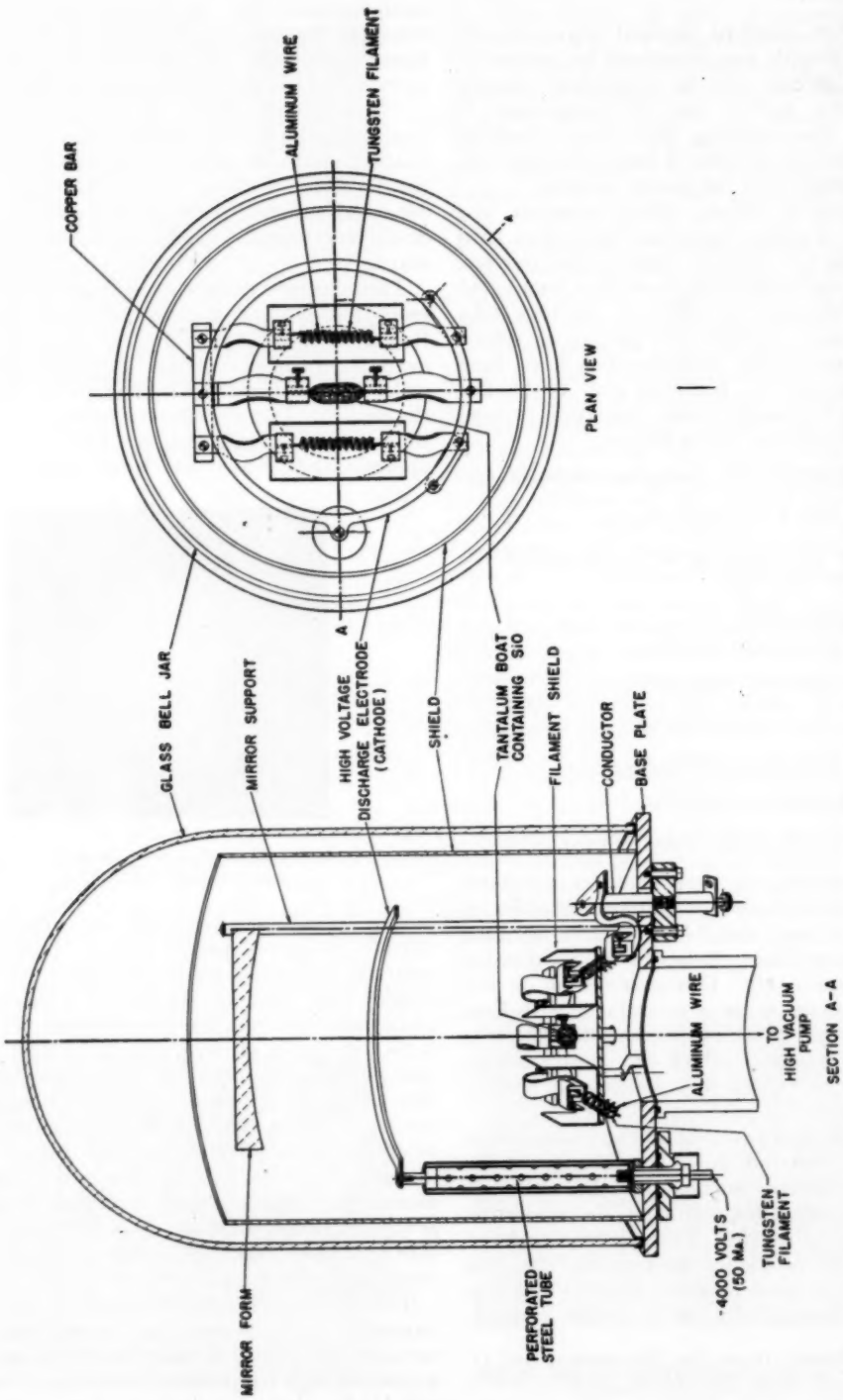


FIG. 5. Evaporator for experimental production of front-surface mirrors protected by silicon monoxide.

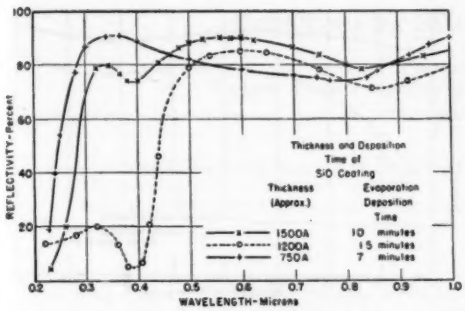


FIG. 6. Reflectivity of SiO-protected aluminum mirrors as a function of wave-length, and variation in reflectivity with film thickness and speed of evaporation. A SiO film thickness of approximately 1500A results in maximum reflectivity. Fast evaporation reduces reflectivity in the ultraviolet and the shorter wave-length portion of the visible.

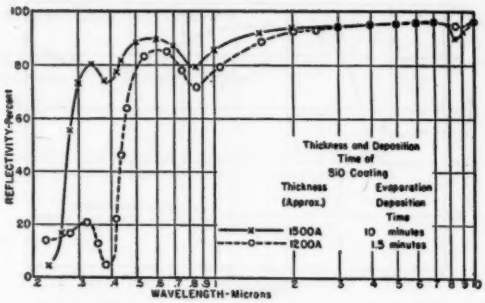


FIG. 7. Reflectivity of SiO-protected aluminum mirrors as a function of wave-length from .2μ to 10μ, and variation in reflectivity with speed of evaporation of SiO.

III. PRODUCTION OF SiO PROTECTED FRONT-SURFACE MIRRORS

Silicon monoxide protected front-surface mirrors may be produced with standard commercial vacuum coating equipment. However, some modifications of the electrodes may be necessary in order to accommodate the high current required by the silicon monoxide evaporation source. The evaporator designed and assembled at the Engineer Research and Development Laboratories for the experimental production of front-surface mirrors protected by silicon monoxide is illustrated in Fig. 5. The vacuum chamber is a glass bell jar, approximately 17 inches in diameter and 27 inches high, placed on a metal base plate. Three evaporation elements are mounted near the base plate; two are tungsten filaments for aluminum evaporation, and the other is a tungsten, tantalum, or molybdenum boat for silicon monoxide evaporation. A shield is placed between each filament and the boat to prevent mutual contamination of the material to be evaporated.

For cleaning the mirror form by glow discharge, an aluminum electrode in the form of a ring is supported horizontally approximately midway between the evaporation sources and the mirror form. This ring is connected to a high voltage, direct current unit for producing a glow discharge at 4000 to 5000 volts, 50 to 100 milliamperes. This high potential electrode is made negative with respect to the base plate which is grounded.

Since much heat is evolved during this discharge cleaning process, the glass bell jar is protected by installing, immediately inside the evaporator, a cylindrical shield of stainless steel, closed at the top. This shield also protects the inside of the bell jar from condensing metal and silicon monoxide during the evaporation process. A hole in the top of this shield permits observation of the mirror form. Two side windows, which can be opened and closed from the outside by a magnet, permit optical determination of the protective film thickness by observing the interference colors on the mirror at grazing incidence.

Silicon monoxide protected front-surface mirrors may be produced on glass, metal, or plastic bases. The techniques for their production are as follows: Glass mirror forms are cleaned with aerosol and concentrated chromic acid cleaning solution and suspended approximately 20 inches above the evaporation sources. The pressure in the chamber is first reduced to about 5×10^{-5} mm Hg to check the vacuum system and then increased to about 3×10^{-3} mm Hg for a 10-minute high voltage glow discharge. The pressure in the chamber is again reduced to about 5×10^{-5} mm Hg and aluminum evaporated onto the mirror form before it has a chance to cool. Immediately thereafter, the protective coating is applied by evaporating silicon monoxide from a tantalum boat. In practice this boat is formed from a piece of sheet tantalum 4 inches by $1\frac{1}{4}$ inches by 0.008 inch. The current required is about 250 amperes. The thickness of the SiO film produced on the aluminum surface is determined by observing the interference colors at an angle near grazing incidence. To obtain highest reflectivity in the visible, the SiO evaporation must be stopped when the interference color, after passing through the colors yellow, red, and green, becomes red the second time. The thickness of the SiO film is then approximately 1500A. The two solid curves of Fig. 6 illustrate the decrease in visible reflection with decrease in film thickness.

The evaporation of the silicon monoxide must be carried out slowly. With the vacuum coating unit described above, five to ten minutes are required to produce optimum SiO films. More rapid evaporation produces yellow films composed of SiO+Si, resulting in undesirable absorption in the blue-violet region. This is illustrated by the broken line reflectance curve of Fig. 6.

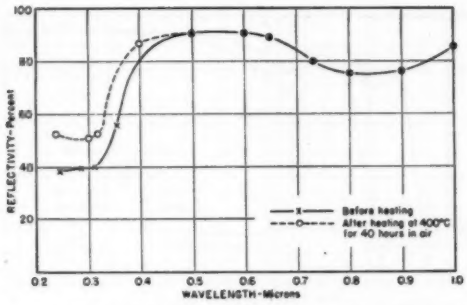


FIG. 8. Reflectivity of SiO-protected aluminum mirror before and after heating in air increases the reflectivity of SiO-protected aluminum mirrors in the ultraviolet without change in the visible.

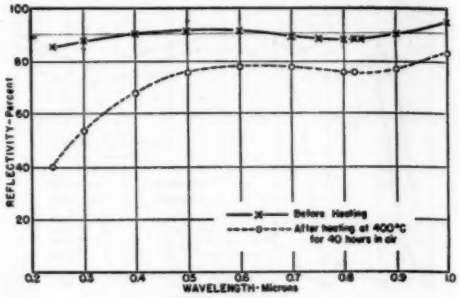


FIG. 9. Reflectivity of unprotected aluminum mirror before and after heating in air. Heating in air substantially reduces the reflectivity of unprotected aluminum mirrors because of oxidation.

To coat metal, a cleaning gaseous discharge is used, employing the metal mirror form as the cathode. Then a SiO film about 2000A thick is evaporated onto the metal form, then aluminum, and, finally, the protective SiO film. The first SiO film improves adherence of the films to the metal form, and prevents the metal of the mirror form from diffusing into the thin aluminum reflecting surface.

IV. REFLECTIVITY OF SiO PROTECTED MIRRORS

The reflectivity in the wave-length region from 0.24 to 10 microns of aluminum mirrors, protected with a slowly and a rapidly evaporated silicon monoxide film of various thickness, is shown in Figs. 6 and 7. Reflectivity of an unprotected aluminum surface, given by the top curve of Fig. 9, reveals that the reflectivity of the protected mirror for the visible is about the same as that for the unprotected mirror. On the other hand, the protected mirror shows high absorption in the ultraviolet and some reflectance decrease by interference in the near infra-red. The absorption in the ultraviolet increases with speed of evaporation of the protective film (see Fig. 6). Figure 7 shows very high reflectivity in the infra-red from approximately 1 micron to 8 microns, with peak reflectivity of about 96 percent. The absorption band, characteristic of silicon oxides, appears between 8 and 10 microns as a small decrease in reflectivity.

V. CHEMICAL AND MECHANICAL RESISTANCE OF SiO-PROTECTED ALUMINUM MIRRORS

SiO-protected aluminum mirrors on glass and metal bases are not adversely affected by prolonged heating in air at temperatures up to 400C, whereas the reflectivity of unprotected aluminum mirrors is substantially reduced by the same treatment. This is illustrated by the comparison of Figs. 8 and 9. Figure 8 also shows that heating in air increases the reflectivity of the protected mirror in the ultraviolet. This results from oxidation of some of the SiO to SiO_2.

The reflectivity of SiO-protected mirrors on glass bases is not changed by boiling in a 5 percent salt solution for a period of 1 hour. These mirrors are also highly resistant to sea water, sodium hydroxide, and hydrochloric acid.

To simulate the most severe cleaning conditions likely to be encountered in service, the abrasion resistance of the mirrors was tested by means of a $\frac{3}{4}$-inch steel ball covered with rough linen, loaded with a 500-gram weight, and subjected to a reciprocating motion. A protected mirror withstands more than 1000 cycles without damage, whereas the reflecting surface of an unprotected aluminum mirror is destroyed in a few cycles.

Conditions for Blackbody Radiation of Gases

WOLFGANG FINKELNBURG
Engineer Research and Development Laboratories, Fort Belvoir, Virginia
(Received August 10, 1948)

A simple derivation is given of the formula for the spectral emission of a thermally excited gas. The accompanying discussion provides a clear answer to the problem of how line emission and blackbody emission of any gas or vapor are interrelated.

IT is a well-known fact that gases, excited by high temperature or any other means, emit essentially discrete line or band spectra characteristic of the subject gas. On the other hand, a cavity radiator emits a continuous spectrum with a spectral energy distribution depending exclusively on the radiator's absolute temperature. This result is usually applied, without much further explanation, to extensive gas masses of equal temperature, and the continuous spectrum of the sun is pointed out as such a case of blackbody radiation of gases. Actually every isothermal gas, under certain conditions, emits a continuous spectrum, the energy distribution of which is given by the Planck formula and thus depends exclusively on its temperature and *not* on its specific nature. As no sufficiently short and clear presentation of the relations between line emission and blackbody emission of gases, though being known basically for a long time, seems to have been published,* it is the purpose of this note to give a simple derivation of the general formula for the spectral emission of thermally excited gases which contains the answer to the question: How is a discrete line spectrum converted into a continuous blackbody spectrum?

Let $J_s(\lambda, T)$ be the spectral emissive power of an ideal blackbody, i.e., the time rate of radiant energy emitted in all directions per cm² of radiating surface. According to the Planck formula, this depends only on the wave-length and the absolute temperature and is measured in watts per cm² per unit wave-length. Let $J(\lambda, T, \cdots)$, on the other hand, be the emissive power of any gaseous radiation source to be calculated, the dots in the brackets indicating that the emissive power depends on additional parameters which will be derived. $A(\lambda, T)$ is the absorptivity of our radiator (the absorbed fraction of any incident radiation) also depending on the wave-length. The emissivity, $E(\lambda, T)$ finally, is the ratio of the emissive power of the source to that of a blackbody at the same temperature (this last being the highest emissive power possible at that temperature):

$$E(\lambda, T) = (J(\lambda, T)/J_s(\lambda, T)). \quad (1)$$

Let us now consider an absorbing layer of d cm thickness with a coefficient of absorption $a(\lambda, p, T)$ per cm, varying with the wave-length and depending on temperature T and pressure p. According to the definition of the absorption coefficient $a(\lambda, p, T)$, the quantity

$$J_d = J_0 \exp[-a(\lambda, p, T)d] \quad (2)$$

emerges from the layer of thickness d where J_0 is the radiation entering the layer. Thus, $J_d/J_0 = \exp[-a(\lambda, p, T)d]$ is the transmitted fraction of the entering radiation. The absorptivity of the layer, i.e., the absorbed fraction of the entering radiation, consequently is

$$A(\lambda, p, T) = 1 - \exp[-a(\lambda, p, T)d]. \quad (3)$$

Now according to Kirchhoff's law we have for a given temperature and pressure

$$J(\lambda)/A(\lambda) = J_s(\lambda), \quad (4)$$

From (4) and (1) we then get immediately the well-known result:

$$E(\lambda) = A(\lambda), \quad (5)$$

that is: The emissivity and the absorptivity of any thermal radiator are equal for given temperature. From (3) and (4) we obtain the spectral emissive power

$$J(\lambda, p, T, d, a) = J_s(\lambda, T)(1 - \exp[-a(\lambda, p, T)d] \quad (6)$$

for a layer of thickness d, absolute temperature T, and absorption coefficient $a(\lambda, p, T)$.** Formula (6) contains the conditions for blackbody radiation of gases: blackbody radiation can be attained asymptotically only as the product of layer thickness times absorption coefficient, $a(\lambda, p, T)d$, approaches infinity for every wave-length.

One evident conclusion from (6) is of importance: If the absorption coefficient $a(\lambda, p, T)$ of any gas or vapor is different from zero for every wave-length, then and only then will the emission approach more and more that of a blackbody of the same temperature with increasing thickness of its radiating layer. This is generally known to be

* W. Finkelnburg, Zeits. f. Physik **88**, 763 (1934).

** In applying this general radiation formula to solid radiators like metal strips, glow bars, etc., the surface reflectivity $R(\lambda, T)$, negligible for gases, has to be taken into account. The right side of (6) then simply has to be multiplied by $(1 - R(\lambda, T))$.

true for continuous radiators, but often has not been realized to be true also for apparently pure line emitters. The physical reason for the transition from the line spectrum to the continuous spectrum with increasing thickness of the radiating layer seems clear. Equation (5) applies not only for the emissivity and absorptivity of the surface, but also for the coefficients of absorption and emission of the gas. In wave-length regions where the coefficient of emission is large (for instance the centers of the lines) the coefficient of absorption is also large, so that only a thin layer of the gas contributes to the radiation (the "optical depth" is small). In regions of small coefficient of emission between the lines, the coefficient of absorption is also small so that for these regions very thick layers contribute appreciably to the observed radiation, i.e., the "optical depth" must be large.

Blackbody radiation thus is approached with increasing thickness of the radiating layer by every gas, if its absorption coefficient is different from zero for every wave-length in the spectral region under consideration. This is always the case, because a thermally excited gas by necessity is ionized to a certain, though occasionally to a very small degree. As a consequence of this ionization, a continuous spectrum resulting from the stopping of the free discharge electrons in the fields of the positive ions covers the whole spectral region.[1] The same applies (with largely varying intensity) for a number of continuous spectra beyond the series limits, where the emission results from recombination of free electrons with ions into different excited states of atoms. Even if any broadening of the discrete lines or bands emitted by the gas is disregarded, the absorption coefficient of every luminous gas thus is different from zero for *all* wave-lengths.

In order to obtain an exact Planck distribution of the emitted spectrum, however, not only an infinite thickness of the radiating layer is necessary according to (6), but also thermal equilibrium must exist throughout the whole gas volume. This latter condition actually never is fulfilled because of the surface cooling caused by the emission of radiation. For both reasons, consequently, blackbody emission can be approached by gases to a much smaller extent than by cavity radiators. The layer thickness of a gas necessary to approach it to a given extent depends exponentially on both the ionization energy, E_i, and the reciprocal of the absolute temperature, $1/T$, and linearly on the reciprocal of the pressure, $1/p$, since the continuous absorption coefficient[2] is given by

$$a = \text{const.} \times p \times e^{-E_i/kT}.$$

The higher the pressure and temperature, and the lower the ionization potential of a gas or vapor are, the thinner thus is the required radiating layer.

[1] W. Finkelnburg, *Kontinuierliche Spektren* (Springer-Verlag, Berlin, 1938). (Reprinted in USA, 1944, by J. W. Edwards, Ann Arbor, Michigan.)

[2] A. Unsoeld, Ann. d. Physik. **33**, 607 (1938).

Symposium on Spectroscopic Light Sources

AN 80-page publication with the above title, in heavy paper cover, can be procured from the headquarters of the American Society for Testing Materials, 1916 Race Street, Philadelphia 3, Pennsylvania, at $2 per copy. Four formal papers are contained in this Symposium as follows: "The Present Status of Excitation in Spectrographic Analysis"—B. F. Scribner, National Bureau of Standards; "A Study of the Controlled Spectrographic Spark Source"—J. H. Enns and R. A. Wolfe, University of Michigan; "Some Properties of Gas Discharges Used as Spectral Sources"—R. C. Mason, Westinghouse Research Laboratories; "Short-Period Phenomena in Light Sources"—G. H. Dieke, The Johns Hopkins University. The pamphlet contains in addition extensive discussions prepared by experts in the field.

The Quantitative Spectrographic Analysis of the Rare Earth Elements*

III. Determination of Major Constituents in Complex Mixtures

VELMER A. FASSEL
Institute for Atomic Research, Iowa State College, Ames, Iowa

(Received November 12, 1948)

Spectrographic methods applicable to the determination of major rare earth constituents in complex mixtures are described. The specific examples discussed are the determination of yttrium and gadolinium in the concentration range of 10 to 100 percent. The procedures involve the high current d.c. arc excitation of rare earth oxide sample-ceric oxide-graphite mixtures. Selected lines of cerium are used as internal standards. The choice of cerium as internal standard was based on the relative ease with which any cerium, if present in a mixture, could be preferentially separated from the mixture and then reintroduced in a standard amount. Precision studies show a standard deviation of about ±2.5 percent for single determinations. The results from a quantitative study of (a) vaporization rates, (b) intensity diagrams of the internal standard line pairs along the longitudinal axis of the arc discharge, and (c) the effect of large changes in excitation conditions on the intensity ratios show that this high precision can be attributed to a high degree of internal standard compensation of excitation variables.

I. INTRODUCTION

IN a previous paper on this subject,[1] spectrographic methods for the determination of small amounts of rare earth impurities in other rare earths were discussed. These procedures were designed to supplement the useful spectrophotometric methods[2,3] by providing a means for determining the rare earths below the concentrations detectable by spectrophotometry. Since several of the rare earths, namely, cerium, lanthanum, yttrium, terbium, and lutecium possess no appreciable absorption in the wave-length regions accessible by ordinary spectrophotometers, it has not been possible to determine even large amounts of these members of the group by spectrophotometric methods. Gadolinium possesses an absorption band in the ultraviolet, but background uncertainty and other interferences limit the general usefulness of this band for analytical purposes. Although spectrographic methods are not generally used for the determination of major constituents, the increasing number of publications[4–8] on these determinations during the last few years indicate that there are no fundamental reasons which prevent these applications. The very limited scope of chemical and other analytical procedures for the analysis of the rare earths especially favored the extension of the spectrographic method to major constituent determinations. In the present paper a method is described for the determination of gadolinium and yttrium in complex mixtures in the concentration range 8 to 100 percent. The method is so designed, however, that it should be adaptable to the determination of any major rare earth constituent except cerium.

II. HISTORICAL

Hopkins and co-workers[9] have investigated the determination of major constituents in rare earth mixtures. Their method was based on the evaporation of rare earth nitrate solutions on graphite electrodes and the excitation of these electrodes with a low voltage d.c. arc. The variations of their results were too large for a successful analytical method; the major error was attributed to the poor reproducibility of the d.c. arc discharge between electrodes prepared in this manner, and to their choice of internal standards.

III. APPARATUS

The spectrograph, optical system, excitation source, electrode assembly, plates, and photographic processing which were used have been described in detail in the previous paper.[1]

IV. PRELIMINARY EXPERIMENTS

The logical experimental approach to the development of methods for the determination of major rare earth constituents was to apply the

* Contribution No. 31 from the Institute for Atomic Research, Iowa State College, Ames, Iowa. Abstracted in part from a thesis presented by Velmer A. Fassel to the Graduate College of Iowa State College for the degree of doctor of philosophy in chemistry. This work was supported by a grant from the Atomic Energy Commission under Contract No. W-7405 eng-82. Presented in part at the meeting of the Optical Society in New York, March 6, 1948.

[1] V. A. Fassel and H. A. Wilhelm, J. Opt. Soc. Am. **38**, 518 (1948).

[2] W. Prandtl and K. Scheiner, Zeits. f. anorg. allgem. Chemie **220**, 107 (1934).

[3] C. J. Rodden, J. Research Nat. Bur. of Stand. **26**, 557 (1941); **28**, 265 (1943).

[4] H. I. Oshry, J. W. Ballard, and H. H. Schrenk, J. Opt. Soc. Am. **32**, 672 (1942).

[5] J. R. Churchill and R. G. Russell, Ind. Eng. Chem. Anal. Ed. **17**, 24 (1945).

[6] E. J. Fitze and W. M. Murray, Ind. Eng. Chem. Anal. Ed. **17**, 145 (1945).

[7] E. K. Jaycox, J. Opt. Soc. Am. **37**, 162 (1947).

[8] A. Kvalheim, J. Opt. Soc. Am. **37**, 585 (1947).

[9] B. S. Hopkins, C. N. McCarty, L. R. Scribner, and M. Lawrenz, Ind. Eng. Chem. Anal. Ed. **10**, 184 (1938).

TABLE I. Operating conditions for obtaining calibration data.

Line pairs	$\frac{\text{Y 4358.726A}}{\text{Ce 4339.317A}}$	Wave-length photographed	4000–4625A
		Order	2nd
	$\frac{\text{Gd 4262.095A}}{\text{Ce 4263.427A}}$	Filter	Corning #774
		Slit	0.04 mm
	$\frac{\text{Gd 4262.095A}}{\text{Ce 4264.370A}}$	Sector apertures	36°, 22.5°, 14.2° (Step sector)
Composition of sample charge	40% CeO_2 10% sample 50% graphite	Duration of excitation	Arced to completion
Wt. of charge	15 mg	Development	4 min. at 21°C Eastman D-19
Type of electrode	Anode: ¼ in. dia. graphite, ¾ in. long with a 2-mm deep cavity and wall thickness of ½ mm. Cathode: ¼ in. dia. graphite 1 in. long and pointed.	Concentration range	8–100%
Excitation source	250 volts d.c. 17 amperes		
Analytical gap	4 mm		
Emulsion	Eastman spectrum analysis #1		

same techniques which were found successful for the determination of minor constituents.[1]

In the analysis of purified rare earths for minor impurities, lines of the major constituent were found to be almost ideal internal standards; however, for complex mixtures the variability in the concentrations of the rare earths precludes the direct use of any of the rare earths present as internal standards. Since tetravalent cerium can be selectively separated from the rest of the rare earths,[10] Hopkins has suggested that any cerium present in a mixture could be removed and then a small, constant amount reintroduced to serve as an internal standard.[9] Hopkins and co-workers abandoned this technique because multiple separations were required completely to remove the cerium. The necessity of making a complete separation can be eliminated by introducing such a large amount of cerium so that the small amount of cerium originally present in a sample, or which remained after a preliminary separation, is rendered negligible. Thus, for mixtures of the type for which this method was designed, no preliminary cerium separation would be required. For other mixtures which contain appreciable cerium, the latter can be separated [10] or determined titrimetrically,[11] and then the proper amount of cerium introduced into the sample to bring its concentration up to the standard amount. The ratio of 80 percent ceric oxide to 20 percent sample was selected for this study.

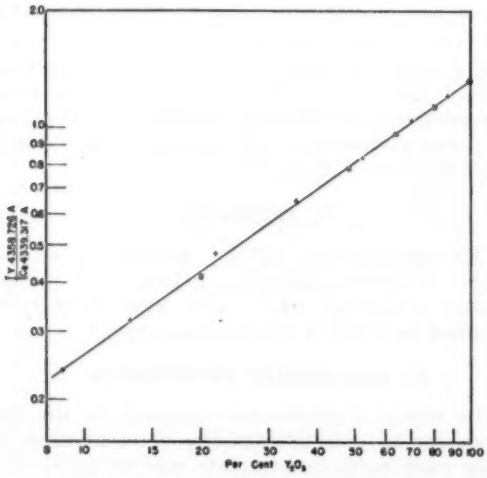


FIG. 1. Calibration curve for the determination of yttrium in complex mixtures.

The addition of a large standard amount of cerium also possessed some desirable secondary features. In the analysis of complex mixtures, the interinfluence of the individual components upon each other during the excitations introduced another variable. The effect of major changes in composition on the discharge temperature, heat conductivity, transport, and distillation rates has

[10] V. A. Chernikhov and T. A. Uspenskaya, Zavodskaya Lab. 9, 276 (1940).
[11] G. F. Smith, *Cerate Oxidimetry* (G. Frederick Smith Chemical Company, Columbus, Ohio, 1942), pp. 48–50.

been shown to be significant.[12] In view of the similarity in the physical properties of the rare earths, this effect was not expected to be significant. Nevertheless, the addition of the relatively large amount of cerium to the sample served to offset the over-all effects of composition changes on the excitation conditions. At the same time, the reduction of the effective concentration of the elements to be determined by the addition of the cerium aided in diminishing the self-reversal effects which generally arise at higher concentrations. Another advantage obtained by the addition of cerium was the reduction in the sample requirements. This is particularly desirable for the analysis of mixtures containing the less abundant members of the series.

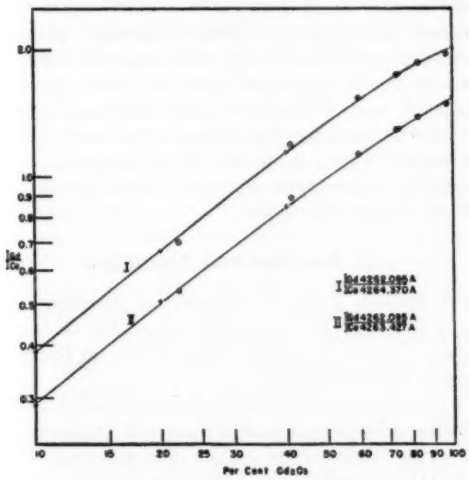


FIG. 2. Calibration curve for the determination of gadolinium in complex mixtures.

V. QUANTITATIVE CALIBRATIONS

A. Preparation of Standards

The initial series of standards was prepared from Hilger's spectroscopically standardized yttrium oxide (Lab. No. 6254) and gadolinium oxide (Lab. No. 13619). After purer fractions became available through the ion-exchange fractionations being carried out in this laboratory,[13] the original calibrations were revised. The calibration curves from each series were consistent with each other after suitable corrections were applied on the purity of the base materials (Figs. 1 and 2). The source of the cerium was primary standard purity ammonium hexanitrato-cerate sold by the G. Frederick Smith Chemical Company, Columbus, Ohio. None of the other rare earths were detected in this compound.

The standards were prepared by combining known volumes of the standard solutions of the rare earths followed by the precipitation, filtration, and ignition of the precipitate to the oxide. The required amount of cerium was added in the form of a standard solution of ammonium hexanitrato-cerate. Since this solution contained tetravalent cerium, it was necessary to reduce the cerium to the trivalent state in order to coprecipitate the cerium and the other rare earths as the oxalates or the hydroxides. This reduction was carried out by acidifying the cerium-rare earth solution with nitric acid and then adding a stoichiometric excess of 30 percent hydrogen peroxide. An orange brown color developed upon the addition of the hydrogen peroxide, but this faded quickly to a colorless trivalent cerium solution during the subsequent heating. The excess hydrogen peroxide was decomposed

TABLE II. Precision of determinations.

Composition of sample	Number of observations	Line pair	Standard percent deviation of intensity ratio for single exposures
39% Gd_2O_3 42% Y_2O_3 19% Other R_2O_3	31	Y 4356.726A / Ce 4339.317A	2.2
39% Gd_2O_3 42% Y_2O_3 19% Other R_2O_3	35	Gd 4262.095A / Ce 4264.370A	1.7
39% Gd_2O_3 42% Y_2O_3 19% Other R_2O_3	32	Gd 4262.095A / Ce 4263.427A	1.9
100% Y_2O_3	16	Y 4356.726A / Ce 4339.317A	3.2
80% Y_2O_3 20% Other R_2O_3	10	Y 4356.726A / Ce 4339.317A	2.3
97% Gd_2O_3 3% Other R_2O_3	9	Gd 4262.095A / Ce 4263.427A	2.7
97% Gd_2O_3 3% Other R_2O_3	9	Gd 4262.095A / Ce 4264.370A	2.1

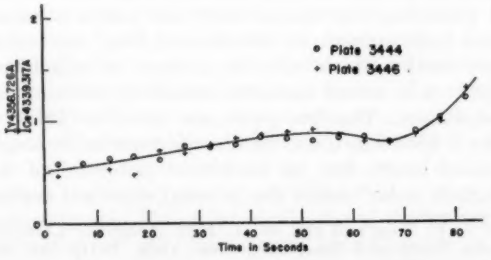


FIG. 3. Variation of analytical intensity ratio with time: yttrium.

[12] For a literature survey on extraneous influence, see E. Fast and J. R. Nielsen, J. Opt. Soc. Am. 37, 614 (1947).

[13] F. H. Spedding, E. I. Fulmer, T. Butler, E. Gladrow, P. Porter, J. Powell, and J. Wright, J. Am. Chem. Soc. 69, 2812 (1947).

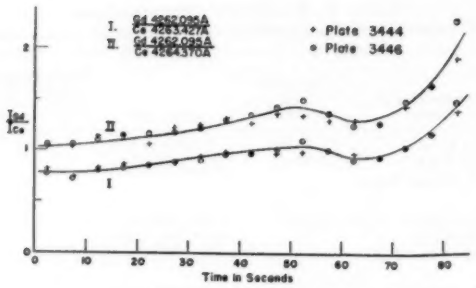


FIG. 4. Variation of analytical intensity ratio with time: gadolinium.

by evaporation of the solution to about 10 ml. After diluting the solution to about 100 ml, a stoichiometric excess of ammonium oxalate was added and this was followed by the addition of ammonium hydroxide until the solution was basic. The formation of a brown precipitate at this point was indicative of incomplete decomposition of the excess hydrogen peroxide. (If a cerous salt is precipitated with ammonium hydroxide in the presence of hydrogen peroxide, a reddish-brown precipitate of perceric hydroxide is formed which is not quantitatively precipitated and is not stable.[14] In the event that the excess hydrogen peroxide was incompletely removed, the solution was again acidified with nitric acid and an excess of 30 percent hydrogen peroxide added to assure reduction of all the cerium oxidized in the basic solution. The solution was then evaporated and treated as outlined above. The rare earth precipitates were filtered off, dried, and ignited to the oxides at 800–900°C. An equal weight of graphite was then mixed with the oxides by grinding for several minutes in an agate mortar. In preparing actual samples for analyses, the proper amounts of sample oxide, ceric oxide, and powdered graphite are mixed by grinding for several minutes in an agate mortar.

B. Selection of Line Pairs

Insofar as possible, line pairs were selected whose components possessed the desired equivalent excitation properties and intensities and proximity of wave-lengths. Special care was taken to select lines least subject to interference from any other rare earths which may be present in significant amounts in actual mixtures containing yttrium and gadolinium. The line pairs are listed in Table I. The Y 4358.726A line, which originates in the singly ionized atom, has an excitation potential of 9.4 electron volts[15] while the internal standard cerium line, which likewise originates in the singly ionized atom, has an excitation potential of 10.6 electron volts.[16] King's[17] temperature classification of the Gd 4262.095A line indicated that it was a blend of two lines in which the violet component originates in one of the higher states in neutral gadolinium while the red component originates in singly ionized gadolinium. The two cerium lines, Ce 4263.427A and Ce 4264.370A, both originate in the ionized atoms with excitation potentials of 10.8 and 10.7 electron volts, respectively.[16]

C. Calibration Experiments

The experimental details used in obtaining the calibration data are summarized in Table I. The photometric evaluations of the spectrograms were made in the usual manner.[1,18,19] The steps of the sectored spectrograms which possessed the optimum degree of blackening were selected for measurement. No background corrections were required. Figures 1 and 2 illustrate the calibration curves obtained by plotting log intensity ratio *vs.* log concentration. The two series of points indicate the substantial agreement between the sets of standards previously discussed.

D. Precision and Accuracy

The data in Table II summarize the results obtained from single repeated exposures of several samples. These exposures were made on from five to seven different plates.

The results obtained from the analysis of various synthetic mixtures of gadolinium and yttrium oxide with other rare earth oxides are summarized in Table III. In addition to substantiating the general accuracy of the calibration curves, these data also indicate that the extraneous influences of large changes in the composition of the mixtures are

TABLE III. Analyses of synthetic mixtures by spectrographic method.

Composition of synthetic mixture	Analytical results: Percent Gd_2O_3 Line pair Gd 4262.095A / Ce 4264.370A	Percent Gd_2O_3 Line pair Gd 4262.095A / Ce 4263.427A
40% Gd_2O_3, 60% La_2O_3	40.0	38.8
50% Gd_2O_3, 50% Sm_2O_3	50.8	51.7
30% Gd_2O_3, 70% Yb_2O_3	31.0	29.5
10% Gd_2O_3, 90% Yb_2O_3	9.8	9.8
	% Y_2O_3	
50% Y_2O_3, 50% La_2O_3	51.0	
50% Y_2O_3, 50% Eu_2O_3	48.5	
25% Y_2O_3, 75% La_2O_3	23.7	
75% Y_2O_3, 25% La_2O_3	76.0	

[14] F. P. Treadwell and W. T. Hall, *Analytical Chemistry* (John Wiley and Sons, Inc., New York, 1931), 9th ed., Vol. 1, p. 567.
[15] W. F. Meggers and H. N. Russell, J. Research Nat. Bur. of Stand. **2**, 733 (1929).
[16] W. E. Albertson and N. H. Hasford, J. Opt. Soc. Am. **31**, 439 (1941).
[17] A. S. King, Astrophys. J. **72**, 221 (1930).
[18] J. R. Churchill, Ind. Eng. Chem. Anal. Ed. **16**, 653 (1944).
[19] W. C. Pierce and N. H. Nachtrieb, Ind. Eng. Chem. Anal. Ed. **13**, 774 (1941).

negligible. The results of some comparative yttrium analyses by the spectrographic and atomic weight methods[20] are summarized in Table IV.

VI. CRITERIA FOR HIGH PRECISION ANALYSES WITH D.C. ARC EXCITATION

The d.c. arc between graphite electrodes possesses definite advantages in convenience, general application, and sensitivity as an excitation source for the spectrographic analysis of inorganic materials, but the poor reproducibility usually associated with this discharge greatly limits its quantitative applications. However, the data in Table II show that d.c. arc excitation can yield results with a precision comparable to that obtained from other carefully controlled excitation sources. The author was aware of several d.c. arc methods in which the externally recognizable features of stability, such as constancy of current and voltage drop across the electrodes, were better controlled than in the procedures described in this paper, whereas the standard deviation was larger by a factor of three to four. This indicated that the high precision in this method may be attributed more to excellent internal standard control of excitation variables than to careful control of external excitation conditions. The following quantitative study on the extent to which the essential criteria of the internal standard were fulfilled verified this conclusion.

Although internal standard techniques are commonly used in d.c. arc excitation methods, it is exceedingly difficult to compensate adequately for several of the excitation variables through the use of internal standards. The source of this difficulty is the thermal mechanism of d.c. arc excitation. This mechanism gives rise to (a) fractional distillation effects which institute large variations in the analytical intensity ratio during the excitation unless internal standards with identical vaporization characteristics are utilized, and (b) large effective excitation temperature variations produced by wandering of the cathode spot, general arc instability, changes in matrix composition, electrode form, spacing, and heat conductivity, and any other process which affects the amount of heat produced in or transferred from the arc. The fact

TABLE IV. Comparative spectrographic and atomic weight method analyses.

Sample number	Percent Y_2O_3 (spectrographic)	Percent Y_2O_3 (atomic weight)
1	92.5	90.5
2	82.5	81.8
3	87.0	84.5
4		

[20] G. L. Barthauer, R. G. Russell, and D. W. Pearce, Ind. Eng. Chem. Anal. Ed. **15**, 548 (1943).

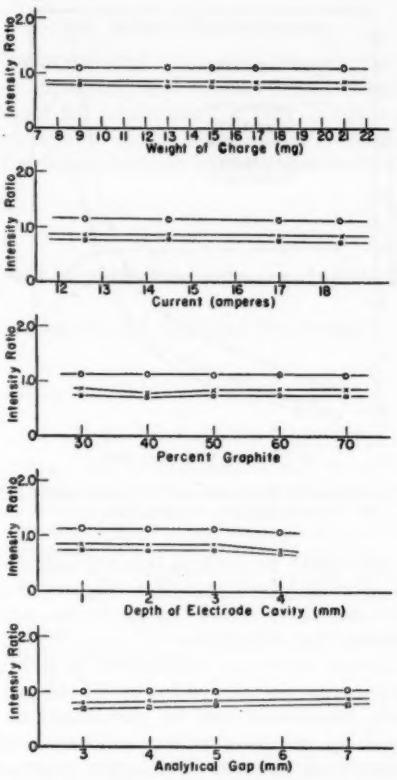

Intensity Ratio
2.0
1.0
0
7 8 9 10 11 12 13 14 15 16 17 18 19 20 21 22
Weight of Charge (mg)
12 13 14 15 16 17 18
Current (amperes)
30 40 50 60 70
Percent Graphite
1 2 3 4
Depth of Electrode Cavity (mm)
3 4 5 6 7
Analytical Gap (mm)

FIG. 5. Effect of changes in excitation conditions on analytical intensity ratios.
○——○ $I_{Gd\ 4262.095A}/I_{Ce\ 4264.370A}$
×——× $I_{Gd\ 4262.095A}/I_{Ce\ 4263.427A}$
⊡——⊡ $I_{Y\ 4358.726A}/I_{Ce\ 4336.317A}$

that the above processes are interrelated makes the excitation process very complicated.

The fractional distillation effects are probably the most difficult to compensate through internal standard control because identity in vaporization response is necessary. For systems which differ in their vaporization properties, the individual line intensities exhibit non-parallel behavior and may even change in opposite directions[21] during the course of the excitation and thus give rise to intensity ratios starting from zero and increasing to progressively higher values (or vice versa). It is apparent that as the over-all changes in intensity ratios increase during the excitations, the reproducibility of the integrated intensity ratios decrease. Moreover, for non-parallel vaporization behavior, the internal standard cannot effectively correct for short period variations in excitation temperatures. The great similarity in the physical properties of the rare earths therefore offers a

[21] L. W. Strock and G. F. Heggen, J. Opt. Soc. Am. **37**, 29 (1947).

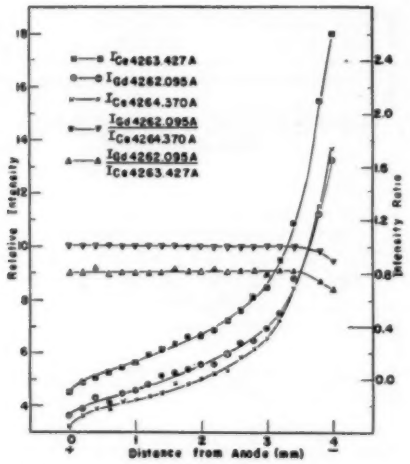


FIG. 6. Intensity diagrams for line pairs used for the determination of gadolinium.

unique possibility to control this variable through internal standard compensation. Figures 3 and 4 show the relatively small change in the intensity ratios during the excitations.

The adverse effects of variations in the excitation temperature occasioned by any of the processes previously mentioned can be minimized or even eliminated by selecting a line pair with similar excitation characteristics. The logical approach to this selection is to utilize the excitation potential data in the spectroscopic literature.[22] The data for the line pairs used in this study were given above. The success with which a line pair minimizes the

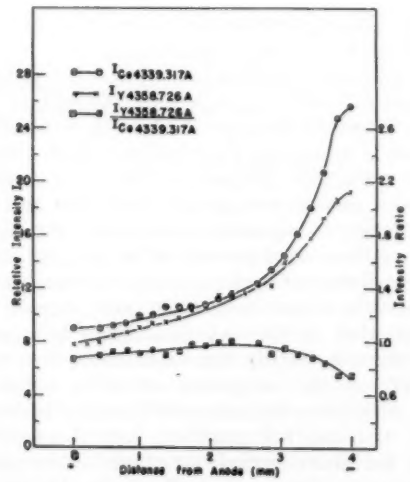


FIG. 7. Intensity diagrams for line pair used for the determination of yttrium.

[22] W. F. Meggers, J. Opt. Soc. Am. 36, 431 (1946).

effects of excitation temperature variations may be determined experimentally by systematically changing the externally controllable variables whose primary or secondary effects produce significant changes in the excitation conditions. Figure 5 summarizes the results from such a study with the line pairs used in this method. Even though large changes in the individual line intensities are instituted by the changes in arc current, weight of charge, and percent graphite, and smaller changes by the other variables, the effect on the intensity ratios is negligible. Besides verifying the homologous nature of the line pairs, these results also indicate the wide range of experimental conditions under which reproducibility is maintained.

Dieke[23] has pointed out that since excitation is not uniform along the length of the arc discharge, the excitation characteristics of a line pair may be compared by observation of the arc column along the longitudinal axis. Most of the voltage drop occurs within the electronic free path of the cathode. Consequently, an enhancement of the spectra which arise in higher energy states and in singly or doubly ionized atoms occurs near the cathode. Therefore, the change in line intensity and intensity ratio under large differences of excitation conditions can be observed by focusing the arc on the slit of a stigmatic spectrograph. Figures 6 and 7 illustrate the diagrams obtained when the spectra were photographed under those conditions. The diagrams clearly show the three- to fourfold parallel increase in the individual line intensities near the cathode, while the intensity ratios show only a strikingly small change. No significant change in the shape of these curves was noted when the samples were arced at 10 amperes rather than the usual 17 amperes. This verified the high degree of current independence of the intensity ratios. In our optical system, the image of the source at the collimating mirror is diaphragmed so that only the central $\frac{2}{3}$ portion of the arc is sampled. Therefore, the slight changes in intensity ratios near the cathode are not recorded on the usual spectrograms. It is apparent that a small change in the optical system, especially if a greater portion of the cathode region is sampled, would greatly increase the individual line intensities, whereas the intensity ratio would undergo only a negligible change. This emphasizes one of the virtues of the internal standard principle when proper line pairs are utilized.

In summary, the data presented above show that the two most important requirements of the internal standard principle, namely, similarity in vaporization and excitation response of the line

[23] G. H. Dieke, War Metallurgy Research Report, Serial No. W-72, *Study of Standard Methods for Spectrographic Analysis.*

pair components, were fulfilled to a high degree in this method. It is realized that the rare earths offer an unusual opportunity to accomplish this. Nevertheless, these results indicate the importance of satisfying these essential criteria if the non-reproducibility of d.c. arc excitation is to be overcome.

VII. ACKNOWLEDGMENTS

The author wishes to express his appreciation to Dr. F. H. Spedding and Messrs. James Wright and Paul Porter for furnishing the pure samples of rare earth oxides used in this study and to Mrs. Mary K. Fassel for her assistance in the experimental work.

JOURNAL OF THE OPTICAL SOCIETY OF AMERICA VOLUME 39, NUMBER 2 FEBRUARY, 1949

The Direct Calculation of Achromatized and Corrected Optical Systems*

ERNESTO E. GEIGER
D. F. Vasconcellos, Optical Instruments Manufacturer, São Paulo, Brazil
Received July 7, 1948

The design of complex lens systems is a tiresome and time-consuming work of successive trigonometric light path calculations, each leading to a new lens bending, air space modification, etc. It can take thousands of man-hours.

A direct process which gives at once the proper lens bending (perhaps the best) and the surfaces' radii, cuts the design work to a small fraction of the classical calculation time. This is the case in this direct process for which a few initial hints are given here.

When the primitive data are related by geometric optics to the entrance and exit pupils, field, aperture, object and image positions, this process leads at once to right results. Generally, they are already good for practical use; when extremely corrected lens systems are needed, it happens that only one partial bending is sufficient to compensate the residual aberrations.

Basis of this new design process is a new equation for achromatization into which enter the lens thicknesses, air spaces, refraction indexes, image position. When data are properly chosen, it is a very rapid job to get the curvature radii of all the optical surfaces.

INTRODUCTION

THE process of optical calculation to be described in this paper is so simple that it is surprising to me that I do not find it mentioned in the classical books on optical design, and I give here a brief description of the method in order that others may be able to profit by it.

In 1942, I was called upon to undertake the design of prismatic binoculars and other instruments for the only optical instrument factory in South America. Being an engineer without experience in optical matters, with no one to turn to for advice, but only books to refer to, I wished to find a direct process of optical calculation to reduce the months of normal ray-tracing to a few hours, and which would be more adapted to the "engineering art" than the conventional methods. I soon found the key, namely, *the deviation of light at the first refracting surface of each lens component must be sensibly equal to the deviation at the last surface.*[1] I verified this practically, and decided to try to make it into a "law" which would underlie the whole process of lens design. The classical conditions for achromatization could no longer be used, so it was necessary to derive new equations for this purpose also. To make a calculation, it is only necessary to know the object and image distances, the lens aperture, and the usual refractive indices of the glasses.

THE FUNDAMENTAL FORMULAS

The formulas to be derived will be applied first to a simple lens, the problem being to calculate the two radii of curvature so that the ray deviation will be the same at each surface. For an achromat, the radius of the cemented surface will be determined for achromatism by a separate formula, and the outer radii will again be chosen to give equal ray deviations at each.**

The sign conventions and symbols will be based on those given by Conrady in his "Applied Optics and Optical Design," the various quantities needed for a single lens being shown in Fig. 1. The object and image points are represented by O and I, the object distance and image distance from the two lens surfaces being, respectively, L_1 and L_2'.

We consider a ray of light entering at a slope-angle U_1 which reaches the lens at the height h above the optical axis, the angle U_1 being negative if the ray is at an anti-clockwise angle from the axis.

* Edited by R. Kingslake.

[1] Steinheil and Voit, *Applied Optics—The Computation of Optical Systems* (Blackie and Son, Ltd., London, 1919), translated by James Weir French, Vol. II, p. 85.

** *Editor's note:* Not all telescope objectives satisfy this condition, but many designs in common use do so remarkably closely. The principle of equal ray deviations at all glass-air surfaces has been frequently employed in condenser design, where it may be used to secure the minimum possible spherical aberration, provided there are no aplanatic surfaces in the system. See, for example, L. T. Sachtleven, J. Soc. Mot. Pict. Eng. 39, 358 (1942).

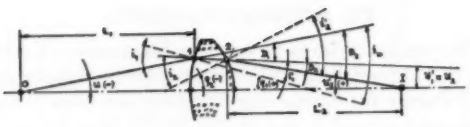


FIG. 1.

The angles of incidence and refraction of this ray at the first surface are i_1 and i_1', respectively, the difference between these angles (i_1-i_1') being the "deviation" (D_1). The normal at the point of incidence forms the angle ρ_1 with the optical axis. The corresponding deviation angle for the second surface is $D_2=i_2-i_2'$, and the slope of the normal at the second point of incidence is ρ_2, which in this case is negative because it is necessary to rotate the axis in a counterclockwise direction to make it coincide with the normal. We consider all lengths positive in the direction of the light, namely, from left to right.

The Total Deviation between the entering and emerging rays is

$$D_t=D_1+D_2$$

To develop the process of calculation, we start with the data of the problem, namely,

(1) The angular semi-aperture (U_2')
(2) The image distance (L_2')
(3) The entrance angle (U_1),

or, alternatively, we may be given the object-to-image distance instead of the image distance (L_2').

In Fig. 2, the path of the ray is shown in outline, with some suitable "zonal thickness" included between the two refracting surfaces, (1) and (2). The point A is the crossing point of the entering and emerging rays. The angle between these two rays is the Total Deviation, $D_t=U'-U$.

For practical purposes, to reduce aberrations, it is advisable that the deviation of a ray at the full aperture of the lens be less than 4°, or better, 3°, at each surface. However, since the deviation at an achromatizing surface is negative, we need to use only one achromat for a total deviation of less than 6°, whereas for a total deviation of 6° to 12°, two achromats should be used in succession. This is the case of oculars, projection lenses, and some photographic lenses.

In the classical treatises, it is explained why it is preferable to make lens calculations for the zonal aperture at 0.7071 of the full semi-aperture of the lens. For our calculations, therefore, U and U' will refer to this zonal aperture.

We will now assign a suitable "zonal thickness," b, to the lens, which might for instance be 4 mm for a lens of 30 mm outside diameter. This b is not in the direction parallel to the optical axis, but it is taken in the direction of the angle $\frac{1}{2}(U'+U)$, or in the direction of the semi-angle of total deviation.

For our purpose, we calculate first the segments a in Fig. 2 (which are equal by our assumption of equal deviations at the two lens surfaces). The projections of a and b on the axis give

$$b\cos\tfrac{1}{2}(U'+U)=a(\cos U+\cos U')$$

whence

$$a=\frac{b\cos\tfrac{1}{2}(U'+U)}{\cos U+\cos U'}.$$

With some experience, we can choose a directly without this calculation.

Referring to Fig. 2, we see that

$$l_1'-L'=a\cos U$$
$$L'-l_2'=a\cos U',$$

or

$$l_1'=a\cos U+L' \tag{1}$$
$$l_2'=L'-a\cos U' \tag{2}$$

hence $b=(l_1'-l_2')\sec\frac{1}{2}(U'+U)$.

However, if the total deviation (D_t) is larger than 6°, we may refer to Fig. 3. Here the meanings of a and b are different. In the case of two achromatized doublets, the measure of b is very close to the zonal thickness of each lens, but a may be 2, 3, or 4 times b, depending on the type of system. Writing $a=k\cdot b$, we have by projection upon the lens axis:

$$l_1'=L'+(k+1)b\cos U \tag{3}$$
$$l_2'=L'+k\cdot b\cos U-b\cos\tfrac{1}{2}(U+U') \tag{4}$$
$$l_3'=L'-k\cdot b\cos U'+b\cos\tfrac{1}{2}(U+U') \tag{5}$$
$$l_4'=L'-(k+1)b\cos U' \tag{6}$$

For the calculations we shall need $(l_1'-l_2')$, $(l_2'-l_3')$, $(l_3'-l_4')$, and l_4'. For systems with a larger number of elements, it will be easy to form equations similar to the above.

If no achromatism is needed, and there are p surfaces $(p/2$ elements), the deviation at each surface is

$$D_1=D_2=D_3=\cdots=D_p=(U'-U)/p=D_t/p,$$

and we get

$$U_1=U$$
$$U_1'=U_2=U+D_1=U+(D_t/p) \tag{7}$$
$$U_2'=U_3=U+2(D_t/p)$$
$$\cdots\cdots\cdots\cdots \text{ etc.}$$

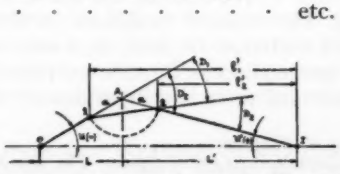


FIG. 2.

CEMENTED SURFACES

To achromatize a lens, a cemented surface must be introduced between the crown and flint glasses, which will contribute some negative deviation. This must be compensated by a higher positive deviation at the two outer or glass-air surfaces. The necessary formulae can be derived from Fig. 4.

The refractive indices on left and right of the cemented surface will be called n_a and n_b, respectively. The light ray crosses the surface at the point A, at height h above the axis. The slope of the normal at the point A is ρ, which is shown negative in Fig. 4.

Considering triangle CAO, we find

$$\rho = i + U, \tag{8}$$

and from the triangle CAI,

$$\rho = i' + U'. \tag{9}$$

Substituting these into the Law of Refraction $n_b/n_a = \sin i/\sin i'$, we get

$$\frac{n_b}{n_a} = \frac{\sin(\rho - U)}{\sin(\rho - U')} = \frac{\sin\rho\cos U - \cos\rho\sin U}{\sin\rho\cos U' - \cos\rho\sin U'}.$$

Dividing by $\cos\rho$, we arrive at

$$\tan\rho = \frac{n_b \sin U' - n_a \sin U}{n_b \cos U' - n_a \cos U}. \tag{10}$$

By this formula, we can calculate the value of ρ at each surface in the system when the slope angles U_1, U_2, etc. are known.

The height of incidence of the ray at the various surfaces can be calculated by

$$\begin{aligned} h_2 &= h_1 - (l_1' - l_2')\tan U_1' \\ h_3 &= h_2 - (l_2' - l_3')\tan U_2'. \\ &\cdots\cdots\cdots \text{etc.} \end{aligned} \tag{11}$$

Then we finally obtain the values of the radii of curvature of the surfaces by

$$\begin{aligned} r_1 &= h_1/\sin\rho_1 \\ r_2 &= h_2/\sin\rho_2 \\ &\cdots \text{etc.} \end{aligned} \tag{12}$$

Another direct method to get the values of the angles (ρ) is the following: As we have already the

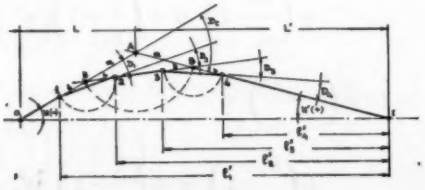


FIG. 3.

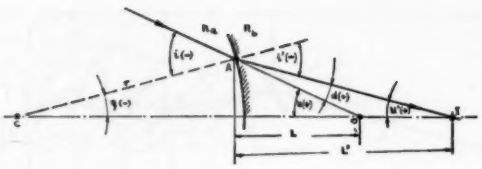


FIG. 4.

values of the partial deviations D_1, D_2, etc., we can calculate the angles of refraction by these approximate relations for small angles:

$$\begin{aligned} i_2' &= \frac{n_a}{n_b - n_a} D_2 \\ i_3' &= \frac{n_b}{1 - n_b} D_3 \\ &\cdots \text{etc.} \end{aligned} \tag{13}$$

and substituting these values, with the slope angles U, in (9) we get the desired values of (ρ). This method is less precise than the previous one.

EQUATIONS FOR ACHROMATISM OF A TELESCOPE OBJECTIVE

The classical formulas for first approximation achromatization do not consider lens thickness, and difficulties are encountered when attempts are made to apply them to the new method of lens calculation. It is, therefore, necessary to develop new formulas for dispersion and achromatization, based upon trigonometric considerations and upon the practical fact that lights of different wavelengths coincide again only at the image point. Furthermore, the new formulas must be prepared especially for each type of lens system, depending on whether it is composed of cemented doublets or of separated lenses. In the present paper, it is intended only to apply some general rules which will be explained in what follows.

To be specific, formulas will be prepared for a cemented doublet followed by prisms, as in the case of a prismatic binocular; formulas for low power microscope objectives and other systems can be similarly derived. These calculations are direct, and it is surprising that the approximate results obtained by their aid are so close to the final solution. For instance, in the example of a binocular objective which will now be given, the first calculation by this method gave only .1397 mm of zonal spherical aberration and .1630 mm of chromatic aberration for the zonal ray, which is already sufficiently close for practical use.

To develop the equations, it would be best to start with the law of refraction, $n_a \sin i = n_b \sin i'$,

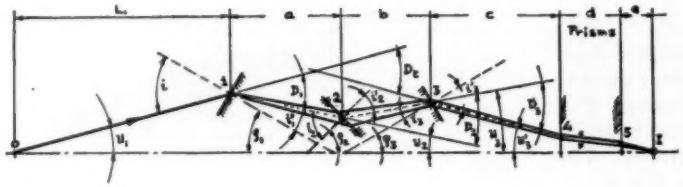


FIG. 5.

but the trigonometry becomes overwhelming. However, since the angles are small, it is near enough to write the angles in place of their sines, which greatly simplifies the problem.

We may now refer to Fig. 5 in which (a) is the axial projection of the zonal thickness of the first lens, generally of crown glass, with index of refraction n_a; (b) is the axial projection of the zonal thickness of the second lens, cemented to the first, made of flint glass, with refractive index n_b; (c) is the axial projection of the zonal thickness of air between objective and prisms; (d) is the total thickness of the prisms, considered as a plano-parallel plate with refractive index n_d, (e) is the air thickness between prisms and image point.

We may now write the standard equation

$$L_1 \tan U_1 + a \tan U_1' + b \tan U_2' + (c+e)\tan U_3' + d \tan U_4' = 0 \quad (14)$$

Since we wish to achromatize the objective, we may write the same equation for the slope-angles (U) of two spectral lines, which for visual achromatism would be C and F. The necessary condition is found by subtracting one of these equations from the other:

$$a(\tan U_{1C}' - \tan U_{1F}') + b(\tan U_{2C}' - \tan U_{2F}') + (c+e)(\tan U_{3C}' - \tan U_{3F}') + d(\tan U_{4C}' - \tan U_{4F}') = 0 \quad (15)$$

Remembering that

$$\tan\alpha - \tan\beta = \frac{\sin(\alpha-\beta)}{\cos\alpha\cos\beta},$$

and admitting that the angles for the sodium D line are intermediate between those for C and F, we get:[2]

$$\frac{a}{\cos^2 U_1'}\sin(U_{1C}' - U_{1F}') + \frac{b}{\cos^2 U_2'} \times \sin(U_{2C}' - U_{2F}') + \frac{(c+e)}{\cos^2 U_3'}\sin(U_{3C}' - U_{3F}') + \frac{d}{\cos^2 U_4'}\sin(U_{4C}' - U_{4F}') = 0 \quad (16)$$

As the differences $(U_C' - U_F')$ are very small, we may replace the sines by the angles themselves. To simplify this, we shall make some substitutions, and obtain an equation in terms of (i_1) (assumed small) which will be the basis of all our calculations. The argument is as follows. For the first surface, we have

$$\begin{cases} U_{1C}' - U_1 = \left(1 - \dfrac{1}{n_a}\right)_C i_1 \\ U_{1F}' - U_1 = \left(1 - \dfrac{1}{n_a}\right)_F i_1. \end{cases} \quad (17)$$

Hence by difference

$$U_{1C}' - U_{1F}' = -\left[\left(\frac{1}{n_a}\right)_C - \left(\frac{1}{n_a}\right)_F\right] i_1. \quad (18)$$

Again, for the second surface,

$$\begin{cases} U_{2C}' - U_1 = \left(1 - \dfrac{1}{n_a}\right)_C i_1 + \left(\dfrac{n_b}{n_a} - 1\right)_C i_2' \\ U_{2F}' - U_1 = \left(1 - \dfrac{1}{n_a}\right)_F i_1 + \left(\dfrac{n_b}{n_a} - 1\right)_F i_2', \end{cases} \quad (19)$$

which gives

$$U_{2C}' - U_{2F}' = -\left[\left(\frac{1}{n_a}\right)_C - \left(\frac{1}{n_a}\right)_F\right] i_1 + \left[\left(\frac{n_b}{n_a}\right)_C - \left(\frac{n_b}{n_a}\right)_F\right] i_2'. \quad (20)$$

The next relation comes from

$$\begin{cases} U_{3C}' - U_1 = \left(1 - \dfrac{1}{n_a}\right)_C i_1 + \left(\dfrac{n_b}{n_a} - 1\right)_C i_2' + \left(\dfrac{1}{n_b} - 1\right)_C i_3' \\ U_{3F}' - U_1 = \left(1 - \dfrac{1}{n_a}\right)_F i_1 + \left(\dfrac{n_b}{n_a} - 1\right)_F i_2' + \left(\dfrac{1}{n_b} - 1\right)_F i_3'. \end{cases} \quad (21)$$

[2] To simplify the formulas, angles and refraction indices referring to sodium line D have no suffix letter.

Since by our hypothesis, the first and third deviations are to be equal,

$$\left(1-\frac{1}{n_a}\right)i_1=\left(\frac{1}{n_b}-1\right)i_3',$$

whence

$$i_3'=-\frac{n_b}{n_a}\cdot\frac{n_a-1}{n_b-1}i_1. \tag{22}$$

Substituting i_3' we get from (21):

$$U_{3C}'-U_{3F}'=-\left[\left(\frac{1}{n_a}\right)_C-\left(\frac{1}{n_a}\right)_F\right]i_1$$
$$+\left[\left(\frac{n_b}{n_a}\right)_C-\left(\frac{n_b}{n_a}\right)_F\right]i_2'-\left[\left(\frac{1}{n_b}\right)_C-\left(\frac{1}{n_b}\right)_F\right]\frac{n_b}{n_a}$$
$$\cdot\frac{n_a-1}{n_b-1}i_1. \tag{23}$$

Finally, our fourth condition is obtained from

$$\begin{cases} U_{4C}'-U_1=\left(1-\frac{1}{n_a}\right)_C i_1+\left(\frac{n_b}{n_a}-1\right)_C i_2' \\ \qquad +\left(\frac{1}{n_b}-1\right)_C i_3'+\left(\frac{1}{n_d}-1\right)_C U_3' \\ U_{4F}'-U_1=\left(1-\frac{1}{n_a}\right)_F i_1+\left(\frac{n_b}{n_a}-1\right)_F i_2' \\ \qquad +\left(\frac{1}{n_b}-1\right)_F i_3'+\left(\frac{1}{n_d}-1\right)_F U_3', \end{cases} \tag{24}$$

which give

$$U_{4C}'-U_{4F}'=-\left[\left(\frac{1}{n_a}\right)_C-\left(\frac{1}{n_a}\right)_F\right]i_1$$
$$+\left[\left(\frac{n_b}{n_a}\right)_C-\left(\frac{n_b}{n_a}\right)_F\right]i_2'-\left[\left(\frac{1}{n_b}\right)_C-\left(\frac{1}{n_b}\right)_F\right]\frac{n_b}{n_a}$$
$$\cdot\frac{n_a-1}{n_b-1}i_1+\left[\left(\frac{1}{n_d}\right)_C-\left(\frac{1}{n_d}\right)_F\right]U_3'. \tag{25}$$

For simplification we now write

$$M=\left(\frac{1}{n_a}\right)_C-\left(\frac{1}{n_a}\right)_F$$

= dispersion of first refracting surface.

$$N=\left(\frac{n_b}{n_a}\right)_C-\left(\frac{n_b}{n_a}\right)_F$$

= dispersion of second refracting surface.

$$O=\left(\frac{1}{n_b}\right)_C-\left(\frac{1}{n_b}\right)_F$$

= dispersion of third refracting surface.

$$P=\left(\frac{1}{n_d}\right)_C-\left(\frac{1}{n_d}\right)_F$$

= dispersion of prisms. (26)

We have now to substitute i_2'. Its value is taken from the following, since the first and third deviations are equal:

$$U_3'-U_1=2\left(1-\frac{1}{n_a}\right)i_1+\left(\frac{n_b}{n_a}-1\right)i_2'.$$

Hence:

$$i_2'=\frac{n_a}{n_b-n_a}(U_3'-U_1)-2\frac{n_a-1}{n_b-n_a}i_1. \tag{27}$$

Substituting this in Eq. (16), we arrive at our desired solution for i_1, namely,

$$\left[\frac{a}{\cos^2U_1'}+\frac{b}{\cos^2U_2'}+\frac{c+e}{\cos^2U_3'}+\frac{d}{\cos^2U_4'}\right]Mi_1$$
$$+2\left[\frac{b}{\cos^2U_2'}+\frac{c+e}{\cos^2U_3'}+\frac{d}{\cos^2U_4'}\right]\frac{n_a-1}{n_b-n_a}Ni_1$$
$$+\left[\frac{c+e}{\cos^2U_3'}+\frac{d}{\cos^2U_4'}\right]\frac{n_b(n_a-1)}{n_a(n_b-1)}Oi_1$$
$$=\left[\frac{b}{\cos^2U_2'}+\frac{c+e}{\cos^2U_3'}+\frac{d}{\cos^2U_4'}\right]\frac{n_a}{n_b-n_a}$$
$$\times N(U_3'-U_1)+\left[\frac{d}{\cos^2U_4'}\right]PU_3'. \tag{28}$$

Having calculated i_1, we obtain at once the first and third deviations:

$$D_1=D_3=[1-(1/n_a)]i_1, \tag{29}$$

or with more precision, the partial deviations may be obtained from

$$D_1=D_3=(i_1-i_1'),$$

the value of i_1' being obtained from the law of refraction

$$\sin i_1'=(n_a/n_b)\sin i_1.$$

The second deviation is found by

$$D_2=(U_3'-U_1)-(D_1+D_3). \tag{30}$$

From these it is easy to calculate

$$U_1'=U_1+D_1=U_2 \tag{31}$$

$$U_2'=U_2+D_2=U_3. \tag{32}$$

The reader will ask how we can calculate Eq. (28) if we do not get U_1' and U_2' until the end. For that equation it is not necessary to have precise values of those angles. The values of a, b, c, d, e in practice vary slightly, as lenses never have the precise thickness for which they were calculated. Also, a and b are very small compared with c, d, and e: in our case, the angle U_3' is a given quantity, and so the only uncertainty lies in the values of U_1' and U_2'. These two angles are almost equal, as the deviation D_2 is generally small. Consequently, for the calculation of Eq. (28), we assume U_1' is equal to U_2', and we determine the value from Eq. (7), referred to in Figs. 2 and 3.

With the correct values, given by Eqs. (31) and (32), we can calculate the radii of curvature by one of the following methods: (a) By use of Eq. (10), we calculate ρ_1, and ρ_2, and ρ_3, and look up $\sin\rho_1$, $\sin\rho_2$, and $\sin\rho_3$. With Eq. (11) we calculate h_2, h_3 and we get the radii by Eq. (12). (b) This method uses Eqs. (8) and (9), whence we obtain rapidly ρ_1, ρ_2, and ρ_3.

EXAMPLE

To explain the procedure more adequately, I shall go through the direct calculation of a very well-known objective used in the standard U. S. Army 6×30 binocular. The available commercial glasses had the following indices:

	n_C	n_D	n_F
(a) Crown lens	1.53828	1.5411	1.54735,
(b) Flint lens	1.61611	1.6210	1.63327,
(d) Prisms	1.56956	1.5725	1.57953.

We assume the zonal thicknesses to be

Crown lens: $a=4$ mm (axial projection),
Flint lens: $b=4$ mm (axial projection),
Focal length: 131.06 mm,
Approximate back focus of objective alone = 126 mm,
Prisms thickness: $d=90$ mm.

Hence the total airspace $(c+e)=69.52179$ mm. Choosing a zonal height of 10.6 mm, we obtain

$$\sin U_3' = \frac{10.6}{131.06} = 0.0808789,$$

whence $U_3' = 4°38'20''.72 = 16{,}700''.72$.

The angle U_1 is, of course, zero.

To start the calculation, we must determine the cosines of the three slopes U_1', U_2', and U_4'. This we do approximately by using Eq. (7), and admitting that U_1' and U_2' are both about equal to $\frac{1}{2}U_3' = 2°19'10''.36$. Hence $\cos U_1' \cong \cos U_2' \cong 0.9991807$; and $\cos^2 U_1' = 0.998362$; $\cos U_3' \cong 0.9967238$; and $\cos^2 U_3' = 0.9934587$. Admitting to first approximation that $U_4' = 2°56'53''.59$, we get $\cos U_4' = 0.9986765$ and $\cos^2 U_4' = 0.9973545$. This approximation is justified because the lens thicknesses are somewhat variable, especially in mass production.

We calculate now the coefficients of Eq. (28):

$$\frac{a}{\cos^2 U_1'} \cong \frac{b}{\cos^2 U_2'} = 4.0065 \text{ mm},$$

$$\frac{c+e}{\cos^2 U_3'} = 69.97956 \text{ mm},$$

$$\frac{d}{\cos^2 U_4'} = 90.2387 \text{ mm}.$$

Also,

$$M = (1/n_a)_C - (1/n_a)_F = 0.003811,$$
$$N = (n_b/n_a)_C - (n_b/n_a)_F = -0.004932,$$
$$O = (1/n_b)_C - (1/n_b)_F = 0.006501,$$
$$P = (1/n_d)_C - (1/n_d)_F = 0.004022$$

and the other coefficients, referred to as the D-line indices, are

$$(n_a - 1/n_b - n_a) = 6.77222,$$
$$(n_a/n_b - n_a) = 19.28786,$$
$$[(n_b(n_a-1)/n_a(n_b-1))] = 0.916512.$$

Substituting these values in Eq. (28), we get

$$i_1 = 27{,}184''.24 = 7°33'4''.24.$$

To calculate the deviations, we calculate the coefficients referred to n_D:

$$(1/n_a)_D = 0.648887; \quad [1-(1/n_a)]_D = 0.351113,$$

obtaining

$$D_1 = D_3 = [1-(1/n_a)]_D i_1 = 9{,}544''.74,$$

and from Eq. (30) we get

$$D_2 = -2{,}388''.76.$$

By Eqs. (31) and (32) we obtain

$$U_1' = 9{,}544''.74 = 2°39'4''.74$$
$$U_2' = 7{,}155''.98 = 1°59'15''.98.$$

Thus our first approximation in determining the coefficients was not wrong, because the difference between U_1' and U_2' was only $39'48''.76$.

To calculate Eqs. (8) and (9), we have to determine i_2' and i_3'. By substitution in Eq. (13) we get:

$$i_2' = 19.28786 \times (-2388''.76) = -46{,}074''.14 = -12°47'54''.14.$$

In the same way $n_b/(1-n_b) = -2.6103$ giving

$$i_3' = -2.6103 \times 9544''.74 = -24{,}914''.69 = -6°55'14''.69.$$

The first angle ρ_1 is equal to i_1 because the entering ray is parallel to the axis. The first zonal height is $h_1 = 10.6$, and from this we get

$$r_1 = h_1/(\sin\rho_1) = +80.6655 \text{ mm}.$$

The second angle ρ_2 is

$$\rho_2 = i_2' + U_2',$$

and since i_2' is negative,

$$\rho_2 = -10°48'38''.16.$$

As

$$h_2 = h_1 - a \tan U_1'$$
$$= 10.6 - 4 \tan 2°39'4''.74 = 10.4152 \text{ mm},$$

we have

$$r_2 = h_2/(\sin\rho_2) = -55.529 \text{ mm}.$$

By the same process,

$$\rho_3 = i_3' + U_3' = -2°16'53''.97.$$

The zonal height is

$$h_3 = h_2 - b \tan U_2' = 10.2764 \text{ mm}.$$

Whence

$$r_3 = h_3/(\sin\rho_3) = -258.1236 \text{ mm}.$$

To calculate the axial thicknesses of the lenses we use the known formula

$$S = r - (r^2 - h^2)^{\frac{1}{2}} \text{ or else } S = r - r \cdot \cos\rho,$$

whence vertex thickness of crown $d_a = 5.68$ mm and flint $d_b = 3.22$ mm. By a direct trigonometrical ray-trace using ordinary methods, these radii and thickness data give the following back-focus distances for the objective alone, and also when the prisms are placed in contact with the rear face of the objective:

	Objective alone	With prisms
Line F, paraxial	126.2103	69.3438
Line C, zonal	126.5964	69.3671
Line F, zonal	126.0703	69.2041

Hence the zonal spherical aberration of the objective alone in F light is only 0.140 mm and the zonal chromatic aberration, F to C, is 0.526 mm, both in the sense of undercorrection. When the prisms are included, the aberrations become

Zonal spherical $(F) = 0.1397$ undercorrected,
Zonal chromatic $(F-C) = 0.1630$ undercorrected.

Nevertheless, these values are good because the aberrations are comparable to the optical tolerances calculated by the Rayleigh formulas.[3]

Focal range $= \lambda/n' \sin^2 U_m'$
Max. marginal spherical aberration $= 4\lambda/n' \sin^2 U_m'$
Max. zonal spherical aberration $= 6\lambda/n' \sin^2 U_m'$
Longitudinal zonal chromatic aberration $= \frac{1}{2}\lambda/n' \sin^2 U_s'$.

As our system operates in air, $n' = 1$, we have

Marginal spherical tolerance $= 0.149$ mm
Zonal spherical tolerance $= 0.223$ mm
Chromatic tolerance $= 0.038$ mm.

Thus our example has somewhat too much chromatic aberration, but the spherical aberration is entirely satisfactory. The residuals can now be eliminated by conventional differential bending formulas explained in the Conrady textbook.

When the calculation is started obtaining (i_1') by the basic condition

$$\sin i'_1 = \frac{n_a}{n_b} \sin i_1$$

and employing Eq. (10), the results are more precise and different.

For the same objective already calculated, we get in this case $r_2 = -55.1550$ mm and $r_3 = -257.6002$ mm.

The author has already applied the direct calculation process to the calculation of oblique pencils, obtaining rapid and good results for complete systems such as telescopes, projection objectives of the "Petzval" type, Taylor-Cooke photographic objectives, Tessar type objective, small power microscope objective and oculars, magnifying glasses up to 10×, etc.

Similar formulae can be readily derived for other and more elaborate systems, but the process is fundamentally the same as that described.

[3] Reported in A. E. Conrady's *Applied Optics and Optical Design* (Oxford University Press, London, 1929), p. 241, and in D. H. Jacobs' *Fundamentals of Optical Engineering* (McGraw-Hill Book Company, Inc., New York, 1943), p. 445.

Erratum: A Differential Adjustment Method of Refining Optical Systems

[J. Opt. Soc. Am. 38, 935–953 (1948)]

WALDEMAR M. STEMPEL
Stevens Institute of Technology, Hoboken, New Jersey

ON page 947, a factor was omitted in the first term on the right of both forms of Eq. (8e). It should read:

$$1_1E_{\log \sin U'} = \frac{\log \sin U}{\log \sin U'} \frac{\tan U}{\tan U'} 1_1E_{\log \sin U} + \text{etc.}$$

$$= \frac{E_{\log \sin U}}{\log \sin U'} \frac{\tan U}{\tan U'} + \text{etc.}$$

Letters to the Editor

Radiant Heat Transfer to Receivers Moving along the Axis of a Right Circular Cylindrical Chamber

S. S. PENNER
Jet Propulsion Laboratory, California Institute of Technology, Pasadena, California
December 3, 1948

AN approximate calculation of radiant energy incident on receivers moving along the axis of a right circular cylinder has been presented previously.[1] When the approximate treatment is replaced by a rigorous analysis where the factor $[1-\exp(-k\rho L)]$ is weighted appropriately with respect to the solid angle and, furthermore, allowance is made for the fact that the radiant energy incident on the receivers is proportional to the apparent area of the bounding surfaces seen from the receivers,[2] then the treatment given previously can be shown to reduce to the following relation:

$$I(t)=(I^0/2)[\epsilon_1(t)+\epsilon_2(t)]. \quad (1)$$

Here $I(t)$ is the total radiant energy incident on unit area of the receiver at any time t, I^0 is the radiant energy emitted over an angle of 2π steradians per unit area of emitter in unit time by a blackbody at the temperature of the radiating gases, and $\epsilon_1(t)$ and $\epsilon_2(t)$ are given by the following expressions, respectively:

$$\epsilon_1(t)=1-\Phi[k\rho vt]-\Phi[k\rho a]+\Phi[k\rho(a^2+(vt)^2)^{\frac{1}{2}}], \quad (2)$$

$$\epsilon_2(t)=1-\Phi[k\rho(L-vt)]-\Phi[k\rho a]+\Phi[k\rho(a^2+(L-vt)^2)^{\frac{1}{2}}], \quad (2a)$$

where $\Phi(z)=\exp(-z)-z\exp(-z)-z^2\int_z^\infty[\exp(-x)/x]dx$, k = average mass emissivity over all wave-lengths, ρ = density of the emitting gases, L = length of cylinder, v = constant linear velocity of the receivers of radiant energy moving along the cylinder axis, and a = radius of cylinder.

Equations (2) and (2a) are equivalent to the relations obtained by Schmidt a number of years ago for a point at the center of the base of a right circular cylinder.[3] It is convenient to define $\epsilon(t)$ through the relation

$$\epsilon(t)=(\tfrac{1}{2})[\epsilon_1(t)+\epsilon_2(t)],$$

whence

$$\epsilon(t)=1-(\tfrac{1}{2})\Phi[k\rho vt]-(\tfrac{1}{2})\Phi[k\rho(L-vt)]-\Phi[k\rho a]+(\tfrac{1}{2})\Phi[k\rho(a^2+(vt)^2)^{\frac{1}{2}}]+(\tfrac{1}{2})\Phi[k\rho(a^2+(L-vt)^2)^{\frac{1}{2}}]. \quad (3)$$

Reference to Eq. (1) shows that the determination of the total radiant energy I_T incident on the receivers moving along the axis of a cylindrical motor chamber requires the integration of Eq. (3) with respect to time, i.e.,

$$I_T=\int_0^{t_r}I(t)dt=I^0\int_0^{t_r}\epsilon(t)dt=(I^0t_r)(1/t_r)\int_0^{t_r}\epsilon(t)dt, \quad (4)$$

where t_r represents the total residence time of the receivers of radiant energy in the cylindrical chamber.

Schmidt[3] has tabulated the values of $\Phi(z)$ for various values of z. Equation (4) is the correct expression for the total radiant intensity incident on the receivers except for neglect of the radiant energy reflected from the (relatively cool) motor walls.

The quantities $\Phi[k\rho(L-vt)]$, $\Phi[k\rho vt]$, $\Phi[k\rho(a^2+(L-vt)^2)^{\frac{1}{2}}]$, and $\Phi[k\rho(a^2+(vt)^2)^{\frac{1}{2}}]$ have been evaluated as a function of t for $L=30$ cm, $k\rho=0.08$ cm^{-1}, $v=1.15\times10^4$ cm/sec., and $a=7$ cm. A plot of the results obtained in this manner reveals the expected symmetry about the center of the axis of the cylindrical chamber. The effective emissivities $\epsilon(t)$ can be evaluated by numerical calculation from Eq. (3). The results obtained in this manner are plotted in Fig. 1 for the first half of the cylinder only, since the emissivity at any point is symmetrical with respect to the center of the chamber axis. Graphical integration shows that

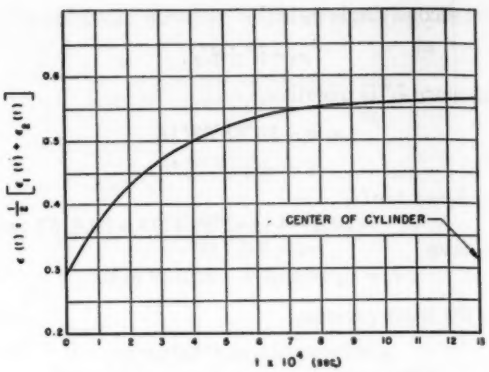


FIG. 1. Variation of effective emissivity $\epsilon(t)$ with time.

$$(1/t_r)\int_0^{t_r}\epsilon(t)dt=0.51,$$

compared with the value 0.63 obtained by the approximate procedure described previously.[1]

The total radiant intensity I_T incident on unit surface area of the receivers of radiant energy is found to be 0.153 cal./sq. cm if $I^0=115$ cal./sq. cm/sec. (corresponding to emitters at a temperature of 3000°K). The corresponding temperature rise ΔT of gaseous absorbers is

$$\Delta T=4I_Tk/c, \quad (5)$$

where c is the heat capacity of the absorbers and the absorption coefficient of the receivers has been assumed to be equal to $k\rho$. From Eq. (5), ΔT is found to be about 50°K.

It may be of interest to note that the preceding laborious calculation leads to a result which differs only by about six percent from the empirical corrected mean hydraulic radius method of Hottel and Port.[4] Thus the corrected mean hydraulic radius method leads to an effective path length L' given by

$$L'=0.85\ (\text{volume of cylinder/surface area of cylinder})$$

or $L'=1.7aL/(a+L)=9.65$ for $a=7$ cm and $L=30$ cm. Hence, $\epsilon=1-\exp(-k\rho L')=0.54$, which is in good agreement with the result 0.51 obtained by the procedure outlined above. In view of the fact that the emissivity $\epsilon(t)$ varies appreciably with time (cf. Fig. 1), it appears likely that the observed excellent agreement between calculated results is to some extent fortuitous and that, for example, the corrected mean hydraulic radius method may lead to erroneous results if the receivers of radiant energy do not move with constant linear velocity.

[1] S. S. Penner and S. Weinbaum, J. Opt. Soc. Am. **38**, 599, 840 (1948).
[2] W. H. McAdams, *Heat Transmission* (McGraw-Hill Book Company, Inc., New York, 1942), Chapter III by H. C. Hottel.
[3] E. Schmidt, Zeits. VDI **77**, 1162 (1933).
[4] Cf. H. C. Hottel and R. B. Egbert, Trans. A.S.M.E. **63**, 297 (1941).

Rapid Scanning Spectrometer for Oscillographic Recording*

B. W. BULLOCK AND S. SILVERMAN
Applied Physics Laboratory, The Johns Hopkins University, Silver Spring, Maryland
November 3, 1948

RECENTLY, there has been considerable interest in the development of rapid methods for scanning spectra.[1,2,3,4] The advent of the PbS, PbSe, PbTe cells

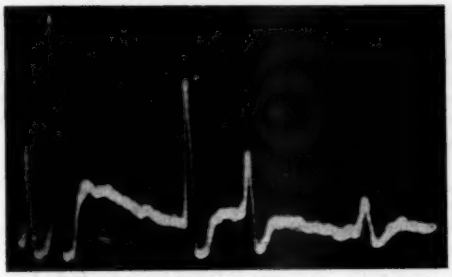


FIG. 1. Mercury arc—visible region—time 0.005 sec.

has increased the attractiveness of rapid scanning as a means for studying infra-spectra of short duration. The writers have built such a spectrometer, and felt it worth while to issue a preliminary note on the instrument.

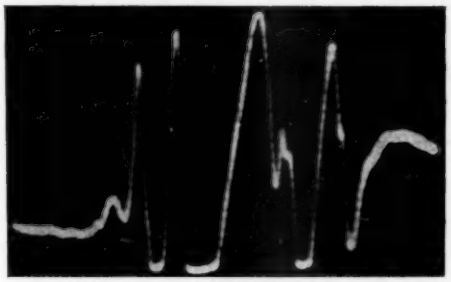


FIG. 2. Gas-oxygen flame—time 0.005 sec.

A conventional Littrow spectrometer with off-axis parabola, aperture *F*.5 was set up with a quartz prism, the exit radiation being focused on a lead sulfide cell (supplied by the Photo MHO Corporation). The Littrow mirror was mounted in a yoke, pivoted on its center line, and connected by a rigid arm to the voice coil of a loudspeaker. When the coil is driven near its resonance frequency (under present load conditions this is 40 cycles per

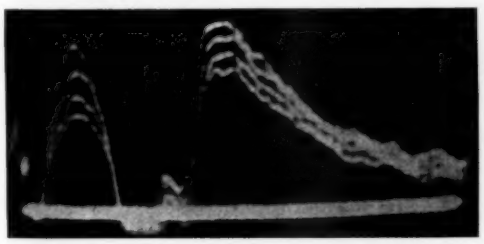


FIG. 3. Four sequences in a carbon monoxide plus oxygen explosion. Time per spectrum =0.005 sec. Time between spectra =0.025 sec.

second) more than sufficient amplitude is available to sweep the entire spectrum from 3600A to 3 microns. Currently, we use approximately ⅓ the available amplitude, in one direction only, to keep the oscilloscope trace clean and to keep $d\theta/dt$ as nearly linear as possible. The photocell output is amplified by a 3-stage wide-band amplifier before being fed to the oscilloscope. The oscilloscope used is a DuMont No. 247, and the single sweep circuit is triggered by the same oscillator that drives the Littrow mirror. Best results are obtained at present with the spectral range limited to approximately 2° of rotation of the mirror; this permits scanning the region 3600A to 3 micron in 3 separate successive intervals. Successful spectra have been obtained at a scanning time of 0.005 sec., repeated at the rate of 40 times per second. Some of these are shown below in Figs. 1, 2, and 3.

The resolving power of this sort of spectrometer is now really determined by the response time of the photo-cell and its amplifier. Since this is in the neighborhood of 2×10^{-5} sec., a scanning time of 2×10^{-3} sec. is about the useful lower limit. Further experiments are in progress, and it is hoped to use an oscillating grating as well as a lead telluride cell to improve both dispersion and spectral range. A more detailed report will be issued in the near future.

* The work described in this paper was supported by the Bureau of Ordnance, U. S. Navy, under Contract NOrd-7386.

[1] F. F. Daly and G. B. B. M. Sutherland, Proc. Phys. Soc. London **59**, 77 (1947).
[2] R. C. Herman and S. Silverman, J. Opt. Soc. Am. **38**, 209 (1948).
[3] W. E. Deal, W. Bradshaw and F. A. Matsen, J. Chem. Phys. **16**, 638 (1948).
[4] S. Silverman, J. Opt. Soc. Am. **38**, 664A (1948).

Light Quanta and Vision

HL. DE VRIES
Rijks Universiteit, Groningen, Holland
December 21, 1948

IN a recent issue of this Journal,[1] a theory was developed on the role of individual light quanta in cone vision. According to the principles developed in the paper, the presence of a test object is already concluded if only one of the cones which are stimulated by the image (or better by its edge) absorbs 4 (in general n) quanta in excess of the quanta originating from the white background (page 750).

This theory leads to a discrepancy which seems to have been overlooked by the authors. A simple analysis shows that for the cases considered, spontaneous fluctuations in the number of quanta, coming from the background, already exceed the value 4. This means that an important part of the cones would be excited even if no test object is present, whereas, according to the theory, one excited cone would already lead to the sensation of "presence of test object." This statement can easily be proved, since the authors themselves considered the fluctuations in the quanta effectively absorbed from the background.

In the first row of Table III we find that the probability of seeing an object is 0.57 if each of 10 cones absorbs (in the mean) 1.62 quanta originating from the object and 68.0 quanta from the background. Simple statistics learns that the probability is about 0.35 that the background gives more than 68.0+3 quanta if in the mean 68.0 are absorbed. Consequently, we find that in the mean 3.5 of the 10 cones absorb 4 quanta in excess of the normal background, whereas only *one* cone needs to receive this number in order to obtain a sensation. The discrepancy becomes even more pronounced for the other rows of Table III.

We can restate the problem as follows. How large must be ν, the number of quanta originating from the background per cone, and n (mentioned above) that the mean number of

spontaneous excitations be small—say 0.1 for 10 cones, or 0.01 per cone.

The second condition is that for those cones which also receive quanta from the object (0.024 ν-quanta in row 1, Table III) the probability of receiving at least n quanta in excess of ν, be 0.57 for the whole set, or about 0.08 for one cone.

By use of a table of the error integral we find that the first condition gives $n \geqq 2.4\nu^{\frac{1}{2}}$. The second condition gives $n - 0.024\nu = 1.4\nu^{\frac{1}{2}}$. So we find $0.024\nu \geqq \nu^{\frac{1}{2}}$ and $\nu \geqq 1700$. Consequently $n \geqq 100$. Of course this result depends on the number of spontaneous sensations, allowed in the first condition. It follows from the experimental arrangement that the value of 0.1 used is still too large. The observer had to decide in which of 8 positions the test object was presented. If the probability is 0.1 that it is seen in one of the seven empty places, the probability is 0.53 that something is *seen* at a wrong place. Consequently the "frequency of seeing right" would never come to 100 percent. Perhaps 0.07 would be a better value. The same calculation then leads to $\nu \geqq 4500$ and $n \geqq 170$. The same reasoning, applied to row 6, Table III (other shape of object), leads to $\nu \geqq 10000$ and $n \geqq 400$. Needless to say that at one fixed brightness n and ν should have one fixed value. (This proves to be not the case in Table III, reference (1), ν varying from 68 to 50 at the same brightness of 2950 foot-lamberts.) A further analysis shows that this third condition leads to $\nu = 10{,}000$ and $n = 400$. There is no *a priori* reason why these large numbers would be impossible. At the brightnesses used, the number of quanta available per cone is large enough.[2] It means, however, that it is no longer true that n is the same at very low brightnesses where it is 4[1] or better two[2] as at very high brightnesses.

Moreover the "probability of seeing" curves calculated from the new formulas prove to become steeper than the experimental curves.

We therefore believe that the leading idea of reference 1, *viz.*, that the detection of the presence of the object is governed by one single cone—is wrong. It seems more probable that the information obtained by all cones is integrated by the nerve system.[2] Perhaps one cone can sometimes give spurious reactions by fluctuations of the background but these reactions need not give rise to a sensation. An indication of discriminating activity of the nerve system has been given already in reference 3; it was found in that paper that *two* cones have to be excited in order to obtain a sensation of light.

[1] E. S. Lamar, S. Hecht, Ch. D. Henley, and S. Shlaer, J. Opt. Soc. Am. **38**, 741 (1948).
[2] Hl. de Vries, Physica **10**, 553 (1943).
[3] M. A. Bouman and H. A. v. d. Velden, J. Opt. Soc. Am. **38**, 570 (1948).

Symposium on Molecular Structure and Spectroscopy

AT the Mendenhall Laboratory of Physics of The Ohio State University from June 13 through June 17, 1949 a Symposium on Molecular Structure and Spectroscopy will be held. Sessions will be devoted to the interpretation of molecular spectroscopic data, the methods of obtaining such data, and to other phases of spectroscopy of current interest. A dormitory will be available for those who wish to reside on the campus during the meeting. Further information, and copies of the program when it becomes available, may be obtained from Professor Harald H. Nielsen, Mendenhall Laboratory of Physics, The Ohio State University, Columbus 10, Ohio.